REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

0802

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE December, 1994	3. REPC Final

4. TITLE AND SUBTITLE
USAF Summer Research Program - 1994 Summer Faculty Research Program Final Reports, Volume 2B, Armstrong Laboratory

5. FUNDING NUMBERS

6. AUTHORS
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research and Development Labs, Culver City, CA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NI
4040 Fairfax Dr, Suite 500
Arlington, VA 22203-1613

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Contract Number: F49620-93-C-0063

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The United States Air Force Summer Faculty Research Program (USAF- SFRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory Technical Directorates and Air Force Air Logistics Centers. Each participant provided a report of their research, and these reports are consolidated into this annual report.

14. SUBJECT TERMS
AIR FORCE RESEARCH, AIR FORCE, ENGINEERING, LABORATORIES, REPORTS, SUMMER, UNIVERSITIES

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1994

SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 2B

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major David Hart

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Gwendolyn Smith

Program Administrator, RDL
Johnetta Thompson

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1994

19981215 117

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Due to its length, Volume 2 is bound in two parts, 2A and 2B. Volume 2A contains #1-22. Volume 2B contains reports #23-45. The Table of Contents for Volume 2 is included in both parts.

This document is one of a set of 16 volumes describing the 1994 AFOSR Summer Research Program. The following volumes comprise the set:

<u>VOLUME</u>	<u>TITLE</u>
1	Program Management Report
	Summer Faculty Research Program (SFRP) Reports
2A & 2B	Armstrong Laboratory
3A & 3B	Phillips Laboratory
4	Rome Laboratory
5A & 5B	Wright Laboratory
6	Arnold Engineering Development Center, Frank J. Seiler Research Laboratory, and Wilford Hall Medical Center
	Graduate Student Research Program (GSRP) Reports
7	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10	Wright Laboratory
11	Arnold Engineering Development Center, Frank J. Seiler Research Laboratory, and Wilford Hall Medical Center
	High School Apprenticeship Program (HSAP) Reports
12A & 12B	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A & 15B	Wright Laboratory
16	Arnold Engineering Development Center

SFRP FINAL REPORT TABLE OF CONTENTS

APPENDICIES:

SFRP FINAL REPORTS

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

x

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

1. INTRODUCTION

The Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), offers paid opportunities for university faculty, graduate students, and high school students to conduct research in U.S. Air Force research laboratories nationwide during the summer.

Introduced by AFOSR in 1978, this innovative program is based on the concept of teaming academic researchers with Air Force scientists in the same disciplines using laboratory facilities and equipment not often available at associates' institutions.

AFOSR also offers its research associates an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. In 1994 the maximum amount of each grant was increased from $20,000 to $25,000, and the number of AFOSR-sponsored grants decreased from 75 to 60. A separate annual report is compiled on the SREP.

The Summer Faculty Research Program (SFRP) is open annually to approximately 150 faculty members with at least two years of teaching and/or research experience in accredited U.S. colleges, universities, or technical institutions. SFRP associates must be either U.S. citizens or permanent residents.

The Graduate Student Research Program (GSRP) is open annually to approximately 100 graduate students holding a bachelor's or a master's degree: GSRP associates must be U.S. citizens enrolled full time at an accredited institution.

The High School Apprentice Program (HSAP) annually selects about 125 high school students located within a twenty mile commuting distance of participating Air Force laboratories.

The numbers of projected summer research participants in each of the three categories are usually increased through direct sponsorship by participating laboratories.

AFOSR's SRP has well served its objectives of building critical links between Air Force research laboratories and the academic community, opening avenues of communications and forging new research relationships between Air Force and academic technical experts in areas of national interest; and strengthening the nation's efforts to sustain careers in science and engineering. The success of the SRP can be gauged from its growth from inception (see Table 1) and from the favorable responses the 1994 participants expressed in end-of-tour SRP evaluations (Appendix B).

AFOSR contracts for administration of the SRP by civilian contractors. The contract was first awarded to Research & Development Laboratories (RDL) in September 1990. After completion of the 1990 contract, RDL won the recompetition for the basic year and four 1-year options.

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM

The SRP began with faculty associates in 1979; graduate students were added in 1982 and high school students in 1986. The following table shows the number of associates in the program each year.

Table 1: SRP Participation, by Year

YEAR	Number of Participants			TOTAL
	SFRP	GSRP	HSAP	
1979	70			70
1980	87			87
1981	87			87
1982	91	17		108
1983	101	53		154
1984	152	84		236
1985	154	92		246
1986	158	100	42	300
1987	159	101	73	333
1988	153	107	101	361
1989	168	102	103	373
1990	165	121	132	418
1991	170	142	132	444
1992	185	121	159	464
1993	187	117	136	440
1994	192	117	133	442

Beginning in 1993, due to budget cuts, some of the laboratories weren't able to afford to fund as many associates as in previous years; in one case a laboratory did not fund any additional associates. However, the table shows that, overall, the number of participating associates increased this year because two laboratories funded more associates than they had in previous years.

3.   RECRUITING AND SELECTION

The SRP is conducted on a nationally advertised and competitive-selection basis. The advertising for faculty and graduate students consisted primarily of the mailing of 8,000 44-page SRP brochures to chairpersons of departments relevant to AFOSR research and to administrators of grants in accredited universities, colleges, and technical institutions. Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs) were included. Brochures also went to all participating USAF laboratories, the previous year's participants, and numerous (over 600 annually) individual requesters.

Due to a delay in awarding the new contract, RDL was not able to place advertisements in any of the following publications in which the SRP is normally advertised: *Black Issues in Higher Education, Chemical & Engineering News, IEEE Spectrum* and *Physics Today*.

High school applicants can participate only in laboratories located no more than 20 miles from their residence. Tailored brochures on the HSAP were sent to the head counselors of 180 high schools in the vicinity of participating laboratories, with instructions for publicizing the program in their schools. High school students selected to serve at Wright Laboratory's Armament Directorate (Eglin Air Force Base, Florida) serve eleven weeks as opposed to the eight weeks normally worked by high school students at all other participating laboratories.

Each SFRP or GSRP applicant is given a first, second, and third choice of laboratory. High school students who have more than one laboratory or directorate near their homes are also given first, second, and third choices.

Laboratories make their selections and prioritize their nominees. AFOSR then determines the number to be funded at each laboratory and approves laboratories' selections.

Subsequently, laboratories use their own funds to sponsor additional candidates. Some selectees do not accept the appointment, so alternate candidates are chosen. This multi-step selection procedure results in some candidates being notified of their acceptance after scheduled deadlines. The total applicants and participants for 1994 are shown in this table.

Table 2: 1994 Applicants and Participants

PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	600	192	30
(HBCU/MI)	(90)	(16)	(7)
GSRP	322	117	11
(HBCU/MI)	(11)	(6)	(0)
HSAP	562	133	14
TOTAL	1484	442	55

4. SITE VISITS

During June and July of 1994, representatives of both AFOSR/NI and RDL visited each participating laboratory to provide briefings, answer questions, and resolve problems for both laboratory personnel and participants. The objective was to ensure that the SRP would be as constructive as possible for all participants. Both SRP participants and RDL representatives found these visits beneficial. At many of the laboratories, this was the only opportunity for all participants to meet at one time to share their experiences and exchange ideas.

5. HISTORICALLY BLACK COLLEGES AND UNIVERSITIES AND MINORITY INSTITUTIONS (HBCU/MIs)

In previous years, an RDL program representative visited from seven to ten different HBCU/MIs to promote interest in the SRP among the faculty and graduate students. Due to the late contract award date (January 1994) no time was available to visit HBCU/MIs this past year.

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates. This year, seven HBCU/MI SFRPs declined after they were selected. The following table records HBCU/MI participation in this program.

Table 3: SRP HBCU/MI Participation, by Year

YEAR	SFRP		GSRP	
	Applicants	Participants	Applicants	Participants
1985	76	23	15	11
1986	70	18	20	10
1987	82	32	32	10
1988	53	17	23	14
1989	39	15	13	4
1990	43	14	17	3
1991	42	13	8	5
1992	70	13	9	5
1993	60	13	6	2
1994	90	16	11	6

6.   SRP FUNDING SOURCES

Funding sources for the 1994 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1994 SRP selected participants are shown here.

Table 4: 1994 SRP Associate Funding

FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	150	98[*1]	121[*2]
USAF Laboratory Funds	37	19	12
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	5	0	0
TOTAL	192	117	133

*1 - 100 were selected, but two canceled too late to be replaced.
*2 - 125 were selected, but four canceled too late to be replaced.

7.   COMPENSATION FOR PARTICIPANTS

Compensation for SRP participants, per five-day work week, is shown in this table.

Table 5: 1994 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994
Faculty Members	$690	$718	$740	$740
Graduate Student (Master's Degree)	$425	$442	$455	$455
Graduate Student (Bachelor's Degree)	$365	$380	$391	$391
High School Student (First Year)	$200	$200	$200	$200
High School Student (Subsequent Years)	$240	$240	$240	$240

The program also offered associates whose homes were more than 50 miles from the laboratory an expense allowance (seven days per week) of $50/day for faculty and $37/day for graduate students. Transportation to the laboratory at the beginning of their tour and back to their home destinations at the end was also reimbursed for these participants. Of the combined SFRP and GSRP associates, 58% (178 out of 309) claimed travel reimbursements at an average round-trip cost of $860.

Faculty members were encouraged to visit their laboratories before their summer tour began. All costs of these orientation visits were reimbursed. Forty-one percent (78 out of 192) of faculty associates took orientation trips at an average cost of $498. Many faculty associates noted on their evaluation forms that due to the late notice of acceptance into the 1994 SRP (caused by the late award in January 1994 of the contract) there wasn't enough time to attend an orientation visit prior to their tour start date. In 1993, 58 % of SFRP associates took orientation visits at an average cost of $685.

Program participants submitted biweekly vouchers countersigned by their laboratory research focal point, and RDL issued paychecks so as to arrive in associates' hands two weeks later.

HSAP program participants were considered actual RDL employees, and their respective state and federal income tax and Social Security were withheld from their paychecks. By the nature of their independent research, SFRP and GSRP program participants were considered to be consultants or independent contractors. As such, SFRP and GSRP associates were responsible for their own income taxes, Social Security, and insurance.

8. CONTENTS OF THE 1994 REPORT

The complete set of reports for the 1994 SRP includes this program management report augmented by fifteen volumes of final research reports by the 1994 associates as indicated below:

Table 6: 1994 SRP Final Report Volume Assignments

	VOLUME		
LABORATORY	SFRP	GSRP	HSAP
Armstrong	2	7	12
Phillips	3	8	13
Rome	4	9	14
Wright	5A, 5B	10	15
AEDC, FJSRL, WHMC	6	11	16

AEDC = Arnold Engineering Development Center
FJSRL = Frank J. Seiler Research Laboratory
WHMC = Wilford Hall Medical Center

6

APPENDIX A -- PROGRAM STATISTICAL SUMMARY

A. Colleges/Universities Represented

Selected SFRP and GSRP associates represent 158 different colleges, universities, and institutions.

B. States Represented

SFRP - Applicants came from 46 states plus Washington D.C. and Puerto Rico. Selectees represent 40 states.

GSRP - Applicants came from 46 states and Puerto Rico. Selectees represent 34 states.

HSAP - Applicants came from fifteen states. Selectees represent ten states.

C. Academic Disciplines Represented

The academic disciplines of the combined 192 SFRP associates are as follows:

Electrical Engineering	22.4%
Mechanical Engineering	14.0%
Physics: General, Nuclear & Plasma	12.2%
Chemistry & Chemical Engineering	11.2%
Mathematics & Statistics	8.1%
Psychology	7.0%
Computer Science	6.4%
Aerospace & Aeronautical Engineering	4.8%
Engineering Science	2.7%
Biology & Inorganic Chemistry	2.2%
Physics: Electro-Optics & Photonics	2.2%
Communication	1.6%
Industrial & Civil Engineering	1.6%
Physiology	1.1%
Polymer Science	1.1%
Education	0.5%
Pharmaceutics	0.5%
Veterinary Medicine	0.5%
TOTAL	100%

Table A-1. Total Participants

Number of Participants	
SFRP	192
GSRP	117
HSAP	133
TOTAL	442

Table A-2. Degrees Represented

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	189	0	189
Master's	3	47	50
Bachelor's	0	70	70
TOTAL	192	117	309

Table A-3. SFRP Academic Titles

Academic Titles	
Assistant Professor	74
Associate Professor	63
Professor	44
Instructor	5
Chairman	1
Visiting Professor	1
Visiting Assoc. Prof.	1
Research Associate	3
TOTAL	192

Table A-4. Source of Learning About SRP

SOURCE	SFRP		GSRP	
	Applicants	Selectees	Applicants	Selectees
Applied/participated in prior years	26%	37%	10%	13%
Colleague familiar with SRP	19%	17%	12%	12%
Brochure mailed to institution	32%	18%	19%	12%
Contact with Air Force laboratory	15%	24%	9%	12%
Faculty Advisor (GSRPs Only)	--	--	39%	43%
Other source	8%	4%	11%	8%
TOTAL	100%	100%	100%	100%

Table A-5. Ethnic Background of Applicants and Selectees

	SFRP		GSRP		HSAP	
	Applicants	Selectees	Applicants	Selectees	Applicants	Selectees
American Indian or Native Alaskan	0.2%	0%	1%	0%	0.4%	0%
Asian/Pacific Islander	30%	20%	6%	8%	7%	10%
Black	4%	1.5%	3%	3%	7%	2%
Hispanic	3%	1.9%	4%	4.5%	11%	8%
Caucasian	51%	63%	77%	77%	70%	75%
Preferred not to answer	12%	14%	9%	7%	4%	5%
TOTAL	100%	100%	100%	100%	99%	100%

Table A-6. Percentages of Selectees receiving their 1st, 2nd, or 3rd Choices of Directorate

	1st Choice	2nd Choice	3rd Choice	Other Than Their Choice
SFRP	70%	7%	3%	20%
GSRP	76%	2%	2%	20%

APPENDIX B -- SRP EVALUATION RESPONSES

1. OVERVIEW

Evaluations were completed and returned to RDL by four groups at the completion of the SRP. The number of respondents in each group is shown below.

Table B-1. Total SRP Evaluations Received

Evaluation Group	Responses
SFRP & GSRPs	275
HSAPs	116
USAF Laboratory Focal Points	109
USAF Laboratory HSAP Mentors	54

All groups indicate near-unanimous enthusiasm for the SRP experience.

Typical comments from 1994 SRP associates are:

"[The SRP was an] excellent opportunity to work in state-of-the-art facility with top-notch people."

"[The SRP experience] enabled exposure to interesting scientific application problems; enhancement of knowledge and insight into 'real-world' problems."

"[The SRP] was a great opportunity for resourceful and independent faculty [members] from small colleges to obtain research credentials."

"The laboratory personnel I worked with are tremendous, both personally and scientifically. I cannot emphasize how wonderful they are."

"The one-on-one relationship with my mentor and the hands on research experience improved [my] understanding of physics in addition to improving my library research skills. Very valuable for [both] college and career!"

Typical comments from laboratory focal points and mentors are:

"This program [AFOSR - SFRP] has been a 'God Send' for us. Ties established with summer faculty have proven invaluable."

"Program was excellent from our perspective. So much was accomplished that new options became viable."

"This program managed to get around most of the red tape and 'BS' associated with most Air Force programs. Good Job!"

"Great program for high school students to be introduced to the research environment. Highly educational for others [at laboratory]."

"This is an excellent program to introduce students to technology and give them a feel for [science/engineering] career fields. I view any return benefit to the government to be 'icing on the cake' and have usually benefitted."

The summarized recommendations for program improvement from both associates and laboratory personnel are listed below (Note: basically the same as in previous years.)

A.   Better preparation on the labs' part prior to associates' arrival (i.e., office space, computer assets, clearly defined scope of work).

B.   Laboratory sponsor seminar presentations of work conducted by associates, and/or organized social functions for associates to collectively meet and share SRP experiences.

C.   Laboratory focal points collectively suggest more AFOSR allocated associate positions, so that more people may share in the experience.

D.   Associates collectively suggest higher stipends for SRP associates.

E.   Both HSAP Air Force laboratory mentors and associates would like the summer tour extended from the current 8 weeks to either 10 or 11 weeks; the groups state it takes 4-6 weeks just to get high school students up-to-speed on what's going on at laboratory. (Note: this same arguement was used to raise the faculty and graduate student participation time a few years ago.)

2. 1994 USAF LABORATORY FOCAL POINT (LFP) EVALUATION RESPONSES

The summarized results listed below are from the 109 LFP evaluations received.

1. LFP evaluations received and associate preferences:

Table B-2. Air Force LFP Evaluation Responses (By Type)

		How Many Associates Would You Prefer To Get ?								(% Response)			
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
Lab	Evals Recv'd	0	1	2	3+	0	1	2	3+	0	1	2	3+
AEDC	10	30	50	0	20	50	40	0	10	40	60	0	0
AL	44	34	50	6	9	54	34	12	0	56	31	12	0
FJSRL	3	33	33	33	0	67	33	0	0	33	67	0	0
PL	14	28	43	28	0	57	21	21	0	71	28	0	0
RL	3	33	67	0	0	67	0	33	0	100	0	0	0
WHMC	1	0	0	100	0	0	100	0	0	0	100	0	0
WL	46	15	61	24	0	56	30	13	0	76	17	6	0
Total	121	25%	43%	27%	4%	50%	37%	11%	1%	54%	43%	3%	0%

LFP Evaluation Summary. The summarized repsonses, by laboratory, are listed on the following page. LFPs were asked to rate the following questions on a scale from 1 (below average) to 5 (above average).

2. LFPs involved in SRP associate application evaluation process:
    a. Time available for evaluation of applications:
    b. Adequacy of applications for selection process:
3. Value of orientation trips:
4. Length of research tour:
5   a. Benefits of associate's work to laboratory:
    b. Benefits of associate's work to Air Force:
6.   a. Enhancement of research qualifications for LFP and staff:
    b. Enhancement of research qualifications for SFRP associate:
    c. Enhancement of research qualifications for GSRP associate:
7.   a. Enhancement of knowledge for LFP and staff:
    b. Enhancement of knowledge for SFRP associate:
    c. Enhancement of knowledge for GSRP associate:
8. Value of Air Force and university links:
9. Potential for future collaboration:
10.   a. Your working relationship with SFRP:
    b. Your working relationship with GSRP:
11. Expenditure of your time worthwhile:

12. Quality of program literature for associate:
13.   a. Quality of RDL's communications with you:
       b. Quality of RDL's communications with associates:
14. Overall assessment of SRP:

Laboratory Focal Point Reponses to above questions

	AEDC	*AL*	*FJSRL*	*PL*	*RL*	*WHMC*	*WL*
= Evals Recv'd	10	32	3	14	3	1	46
Question =							
2	90 %	62 %	100 %	64 %	100 %	100 %	83 %
2a	**3.5**	**3.5**	4.7	4.4	4.0	4.0	**3.7**
2b	4.0	3.8	4.0	4.3	4.3	4.0	3.9
3	4.2	3.6	4.3	3.8	4.7	4.0	4.0
4	3.8	3.9	4.0	4.2	4.3	NO ENTRY	4.0
5a	4.1	4.4	4.7	4.9	4.3	3.0	4.6
5b	4.0	4.2	4.7	4.7	4.3	3.0	4.5
6a	**3.6**	4.1	**3.7**	4.5	4.3	**3.0**	4.1
6b	3.6	4.0	4.0	4.4	4.7	3.0	4.2
6c	3.3	4.2	4.0	4.5	4.5	3.0	4.2
7a	3.9	4.3	4.0	4.6	4.0	3.0	4.2
7b	4.1	4.3	4.3	4.6	4.7	3.0	4.3
7c	3.3	4.1	4.5	4.5	4.5	5.0	4.3
8	4.2	4.3	5.0	4.9	4.3	5.0	4.7
9	3.8	4.1	4.7	5.0	4.7	5.0	4.6
10a	4.6	4.5	5.0	4.9	4.7	5.0	4.7
10b	4.3	4.2	5.0	4.3	5.0	5.0	4.5
11	4.1	4.5	4.3	4.9	4.7	4.0	4.4
12	4.1	3.9	4.0	4.4	4.7	3.0	4.1
13a	**3.8**	**2.9**	4.0	4.0	4.7	**3.0**	**3.6**
13b	**3.8**	**2.9**	4.0	4.3	4.7	**3.0**	**3.8**
14	4.5	4.4	5.0	4.9	4.7	4.0	4.5

3. 1994 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 275 SFRP/GSRP evaluations received.

Associates were asked to rate the following questions on a scale from
1 (below average) to 5 (above average)

1.	The match between the laboratories research and your field:	4.6
2.	Your working relationship with your LFP:	4.8
3.	Enhancement of your academic qualifications:	4.4
4.	Enhancement of your research qualifications:	4.5
5.	Lab readiness for you: LFP, task, plan:	4.3
6.	Lab readiness for you: equipment, supplies, facilities:	4.1
7.	Lab resources:	4.3
8.	Lab research and administrative support:	4.5
9.	Adequacy of brochure and associate handbook:	4.3
10.	RDL communications with you:	4.3
11.	Overall payment procedures:	**3.8**
12.	Overall assessment of the SRP:	4.7

13.   a. Would you apply again?   Yes: 85%
           b. Will you continue this or related research?   Yes: 95%

14. Was length of your tour satisfactory?   Yes: 86%

15. Percentage of associates who engaged in:

    a. Seminar presentation: 52%
    b. Technical meetings: 32%
    c. Social functions: 03%
    d. Other   01%

16. Percentage of associates who experienced difficulties in:

    a   Finding housing: 12%
    b. Check Cashing: 03%

17. Where did you stay during your SRP tour?

    a. At Home: 20%
    b. With Friend: 06%
    c. On Local Economy: 47%
    d. Base Quarters: 10%

THIS SECTION FACULTY ONLY:

18. Were graduate students working with you?   Yes: 23%

19. Would you bring graduate students next year?   Yes: 56%

20. Value of orientation visit:

    Essential: 29%
    Convenient: 20%
    Not Worth Cost: 01%
    Not Used: 34%

THIS SECTION GRADUATE STUDENTS ONLY:

21. Who did you work with:

    University Professor: 18%
    Laboratory Scientist: 54%

4. 1994 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

The summarized results listed below are from the 54 mentor evaluations received.

1. Mentor apprentice preferences:

Table B-3. Air Force Mentor Responses

| | | How Many Apprentices Would You Prefer To Get ? | | | |
| | | *HSAP Apprentices Preferred* | | | |
Laboratory	*# Evals Recv'd*	*0*	*1*	*2*	*3+*
AEDC	6	0	100	0	0
AL	17	29	47	6	18
PL	9	22	78	0	0
RL	4	25	75	0	0
WL	18	22	55	17	6
Total	54	20%	71%	5%	5%

Mentors were asked to rate the following questions on a scale from
1 (below average) to 5 (above average)

2. Mentors involved in SRP apprentice application evaluation process:
    a. Time available for evaluation of applications:
    b. Adequacy of applications for selection process:
3. Laboratory's preparation for apprentice:
4. Mentor's preparation for apprentice:
5. Length of research tour:
6. Benefits of apprentice's work to U.S. Air force:
7. Enhancement of academic qualifications for apprentice:
8. Enhancement of research skills for apprentice:
9. Value of U.S. Air Force/high school links:
10. Mentor's working relationship with apprentice:
11. Expenditure of mentor's time worthwhile:
12. Quality of program literature for apprentice:
13.   a. Quality of RDL's communications with mentors:
    b. Quality of RDL's communication with apprentices:
14. Overall assessment of SRP:

Question #	AEDC	AL	PL	RL	WL
# Evals Recv'd	6	17	9	4	18
2	100 %	76 %	56 %	75 %	61 %
2a	4.2	4.0	3.1	3.7	3.5
2b	4.0	4.5	4.0	4.0	3.8
3	4.3	3.8	3.9	3.8	3.8
4	4.5	3.7	3.4	4.2	3.9
5	3.5	4.1	3.1	3.7	3.6
6	4.3	3.9	4.0	4.0	4.2
7	4.0	4.4	4.3	4.2	3.9
8	4.7	4.4	4.4	4.2	4.0
9	4.7	4.2	3.7	4.5	4.0
10	4.7	4.5	4.4	4.5	4.2
11	4.8	4.3	4.0	4.5	4.1
12	4.2	4.1	4.1	4.8	3.4
13a	3.5	3.9	3.7	4.0	3.1
13b	4.0	4.1	3.4	4.0	3.5
14	4.3	4.5	3.8	4.5	4.1

5. 1994 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 116 HSAP evaluations received.

HSAP apprentices were asked to rate the following questions on a scale from
1 (below average) to 5 (above average)

1. Match of lab research to you interest   3.9

2. Apprentices working relationship with their mentor and other lab scientists: 4.6

3. Enhancement of your academic qualifications: 4.4

4. Enhancement of your research qualifications: 4.1

5. Lab readiness for you: mentor, task, work plan   **3.7**

6. Lab readiness for you: equipment supplies facilities   4.3

7. Lab resources: availability   4.3

8. Lab research and administrative support: 4.4

9. Adequacy of RDL's apprentice handbook and administrative materials: 4.0

10. Responsiveness of RDL's communications: **3.5**

11. Overall payment procedures: **3.3**

12. Overall assessment of SRP value to you: 4.5

13. Would you apply again next year?   Yes: 88%

14. Was length of SRP tour satisfactory?   Yes: 78%

15. Percentages of apprentices who engaged in:

    a. Seminar presentation:
    b. Technical meetings: 48%
    c. Social functions: 23%
                                    18%

AN EXPLORATORY STUDY OF WEIGHTED FUZZY KEYWORD BOOLEAN RETRIEVAL WITH HYPERTEXT LINKS FOR THE CASHE:PVS SYSTEM

Donald H. Kraft
Professor
Department of Computer Science

Louisiana State University
Baton Rouge, LA470803-4020

AN EXPLORATORY STUDY OF WEIGHTED FUZZY KEYWORD BOOLEAN RETRIEVAL WITH HYPERTEXT LINKS FOR THE CASHE:PVS SYSTEM

Donald H. Kraft
Professor
Department of Computer Science
Louisiana State University

Abstract

The purpose of this project has been to explore the possibilities of providing improved retrieval and browsing capabilities to the CASHE:PVS system. That system currently provides access to human engineering studies and allows users to navigate from one entry to another and to simulate ergonomic experiments in order to provide understanding and improved design. The use of keywords, weights, and fuzzy Boolean logic was explored last summer in order to demonstrate the feasibility of that approach to improve the retrieval capabilities of CASHE:PVS, based in large part on the series of sample queries constructed for CASHE:PVS and known as the Design Checklist. This summer, demonstration of the feasibility of automatically generating additional hypertext links between EDC entries based on a similar approach involving term frequencies was accomplished. Future steps needed to expand this approach and this methodology and to integrate them into the CASHE:PVS system are presented.

AN EXPLORATORY STUDY OF WEIGHTED FUZZY KEYWORD BOOLEAN RETRIEVAL WITH HYPERTEXT LINKS FOR THE CASHE:PVS SYSTEM

Donald H. Kraft

Introduction

The Computer Aided Systems Human Engineering: Performance Visualization System (CASHE:PVS) has been developed and is being prepared to be marketed for use. CASHE:PVS consists of a multi-media ergonomics database containing the complete Engineering Data Compendium (EDC) and MIL-STD-1472D, plus a unique visualization tool, the Perception and Performance Prototyper (P^3). Through cooperation with NASA and NATO, plus Tri-Services, the EDC was produced to define approaches to communicate human factors data to designers (levels of technical content, presentation format and style, and terminology) and to provide access to specific technical information relevant to design problems. The MIL-STD-1472D is the military standard for human engineering design criteria, including systems, equipment, and facilities. This standard includes over eighty figures and tables, and is used to insure human-systems integration as well as efficiency, reliability, safety, trainability, and maintainability in acquired systems [Boff91a, Boff91b]. CASHE:PVS's goal is to enable "ergonomics to be supported as a 'full partner' among other design disciplines within a computer-aided environment. By fully integrating ergonomics into the systems design process, more effective human-system designs can be visualized."

Just over a decade ago, the Integrated Perceptual Information for Designers (IPID) project was underway to begin to aid the accessibility and use of ergonomics data in design. This included the identification, collection, and consolidation of human performance data; the representation and presentation of such data to designers (human engineering); the training of designers in the use of ergonomics data; and the definition and evaluation of integrated media options to allow designers to access, interpret and apply such data. Out of these efforts has come the CASHE:PVS system.

The heart of the CASHE:PVS system, which resides on an Macintosh II computer, is its reference databases which include the previously mentioned EDC and the MIL-STD-1427D. Individual users can also customize the system, adding annotations (perhaps to the research contained in the EDC entries), storing files, and otherwise augmenting the reference databases. CASHE:PVS also contains P^3, the test bench simulator mentioned above. A user interested in what types of alarms would be best in a noisy environment, such as the cockpit of an airplane, can look up the appropriate entry(entries) in the EDC and peruse it (them). However, looking at charts and tables of data generated by human factors experts may not provide the full understanding of what the numbers mean for a design engineer. Thus the

P^3 allows the user to specify a given level of noise and type of alarm and then hear for him/herself what the numbers "sound like." Moreover, users can construct additional simulated tests to go beyond the entries for better design [Cona93].

A sample of an EDC entry is found in [Boff88a]. There are up to ten sections of an EDC entry. These include a title with a concise description of the entry content (including a number based on a topic outline); key terms to verify entry content and to provide access points to the entry via index search; a general description to summarize the entry content findings, results, conclusions, models, laws, or principles); an applications section to describe areas of application for the entry (e.g., types of displays); a methods section to describe how data was collected; an experimental results section that may contain graphics and tables with details on the results of data analysis; an experimental validation section to show how a model, law, or principle was verified; a section on constraints to show limitations in the application of the entry results (e.g., criteria that must be met for proper application); a set of key references with bibliographical citations to original literature with more detailed information on the topics of the entry; and cross-references to other entries on the topics in the entry [Boff91a, Boff91b].

The user can use the TextViewer component of CASHE:PVS to view the text entry from the EDC, once an entry has been selected. Users can browse through any of the fields mentioned above. Users can also navigate to other EDC entries. In addition, FigureViewer and TableViewer components of CASHE:PVS allow the user to view the entry graphs and tables, respectively [Boff91a,Boff91b]. In addition, as mentioned above, CASHE:PVS users can simulate situations such as noise or vibration to sense directly meaningful situations of the data in action. However, one component needing more capability is the provision of additional navigation aids to enhance a user's ability to access relevant information.

The Weighted (Fuzzy) Retrieval Approach

In effect, CASHE:PVS is a multi-media system that allows a user to search through the EDC and MIL-STD for ergonomic information that will be helpful in design. Users can navigate to some extent from one entry to another via a number of paths, such as an index, full text queries, and cross-references to other entries. There are also links to the appropriate portions of the military standard (MIL-STD-1472D). It is the strengthening of those links that is the focus of this year's summer project. Last year's project [Kraft93] was intended to explore means of employing fuzzy set theory and other information retrieval mechanisms to better to allow a user to identify one or more EDC entries of interest. The basis of that research project was the employment of keywords and weights for those keywords in order to provide an appropriate match between a user request and the EDC collection. Moreover, the focal point was the use of the Design Checklist, a series of queries hierarchically put together as sample user queries.

The EDC entries are organized along a hierarchical outline as presented in the EDC Table of Contents. The numbering system as part of the title field reflects that hierarchical ordering. For example,

consider the EDC entry numbered 1.104; the category 1.0 reflected "visual acquisition of information", and the sub category 1.1 reflected "measurement of light"; hence the entry on "measurement of radiant and luminous energy" fell in this category and sub category.

In order to make CASHE:PVS more useful to designers, a series of queries were generated. These sample queries are there to serve as examples of the kinds of questions designers might ask. Each query consists of one or more questions, such as "For extended light sources to be treated, with small error, as point sources, how much greater than their diameter must their distance be from illuminated surfaces?". Moreover, each query has a list of one or more EDC entry numbers referring to the EDC entry (entries) that could answer the question(s) in the query (e.g., 1.104). There are 1,069 queries in the set, which is known as the Design Checklist. The Design Checklist is ordered hierarchically by subject; however, this hierarchy is independent of the EDC hierarchy. It was decided early on that the best approach for last year's project was to generate a keyword mechanism for access to the Design Checklist queries and then to link the retrieved queries to the EDC entries. It was felt that this same approach might be fruitful for automatic generation of hypertext links.

Information Retrieval

We first generate a model of information retrieval in order to better understand the background from which both last summer's project and this summer project have been developed. Information retrieval, unlike standard database systems, is in general concerned with the imprecise nature of determining which textual records are relevant to users in response to queries. Previous research has involved document representation (e.g., indexing) to determine which terms to use for topicality. Of course, it is well known that relevance itself is an imprecise concept, incorporating many factors besides topicality. Weights specifying a degree of "aboutness", with a fuzzy set interpretation, can be incorporated to induce a document ranking mechanism. Weights on terms in the user query can also be added to specify a degree of relative term importance. Let us consider a retrieval system as a set of records that are identified, acquired, indexed, and stored, plus a set of user queries for information that are matched to the index to determine which subset of the stored records should be retrieved and presented to the user in which order.

We can begin to model the retrieval system by the following [Kraft85,Kraft83,Kraft93]. Consider D to be a set of documents, or textual records, from which we wish to retrieve subsets of documents in response to users. In order to match documents to user requests, we need to describe the contents of the documents in some manner. A most common mechanism for doing this, especially in terms of doing this automatically by computer, is through the use of keywords [Salton89]. Thus, consider T to be a set of index terms, i.e., keywords. It may be the case that some of the members of T are phrases (e.g., "information retrieval"), but we will consider only single-word keywords.

The assignment of keywords to documents is a process often known as indexing. It is a difficult,

complex task, but can be represented algebraically as:

F: the indexing function, where F: D x T ---> {0,1}.

This implies that one takes a given document d and a given term t, indexes d with regard to t through the function F that produces a 1 (t is used to describe d) or a 0 (term t is not used to describe d). Thus, for the EDC entry whose title is "Measurement of Radiant and Luminous Energy," we might use the term "illumination" (an F of 1) but not use the term "auditory" (an F of 0).

However, the notion of weights, i.e., fuzziness, has long been considered not only feasible but desirable [Kraft85]. Thus, we can let the range of F be the continuous unit interval [0,1], so that partial indexing is allowed. Thus, we might decide that the document in question is about the concept "irradiance" 0.85, while it is about "flux" 0.45. That is to say, the indexing function F maps a given document d and a given indexing term t to a number between 0 and 1 (0 implies that the document is not at all about the concept(s) represented by term t and 1 implies that the document is perfectly represented by the concept(s). Thus, we have a fuzzy set with F being the membership function, mapping the degree to which document d belongs in the set of documents "about" the concept(s) represented by term t.

There are several means of estimating F for a given situation. Salton [Salton89] suggests a measure based on the inverted document frequency (IDF), defined as

$$f_{i,t} * \log [N / N_t]$$

where $f_{i,t}$ is the number of times term t occurs in document i, N is the number of documents in the collection (database), and N_t is the number of documents in the collection in which term t occurs at least once. The IDF is the log factor in the formula. We note that Salton's measure gives larger weights to terms that occur frequently in the document in question but do not occur at all in most documents, and furthermore that Salton's weight needs to be normalized to be contained in the interval [0,1]. Moreover, there are other, albeit similar, mechanisms by which to calculate the weights [Kraft94a].

Most retrieval systems employ a stop word list of common words (e.g., "a", "an", "and", "the", "by", "for", "generally" "one", "once", and "worked") that are to be ignored, since these words are too common and generally convey little if any topicality information about documents that contain them. The remaining words are known as terms or keywords. We also note that if one restricts F to the set {0,1}, one has the case of classical indexing, where a specific term is either attached to the document (1) or it is not (0). By allowing F to also take on values in the open interval (0,1), one can weight terms according to their importance or significance in describing the content of a document in order to better retrieve, and rank, it for users who want it. In this latter case, F has been called an index term weight.

Now, let us define Q as the set of user queries for documents. This leads to the need to represent the query with a set of keywords. This is stated algebraically as

a: Q x T ---> {0,1} = a(q,t) = the importance of term t in describing the query q

We note that the function a assigns a term t to the query (a = 1) or it leaves out the term t (a = 0).

However, again we may wish to generalize this by incorporating weights on the terms to indicate relative importance; thus we again let the range become the unit interval [0,1]. Clearly, not all terms are equally important in a given query. In determining which houses to consider from listings in the real estate section of a newspaper, terms describing whether there is a fireplace or not may not be as important as terms describing the cost of the house.

It is here in the processing of the query, i.e., evaluating the documents, one at a time, against the query, that one begins to introduce problems in terms of maintaining the Boolean lattice [Kraft83]. Because of this, certain mathematical properties can be imposed on F, but more directly on a and on the matching procedure [Kraft85]. To describe the matching procedure, consider the function g, defined as

$$g: [0,1] \times [0,1] \longrightarrow [0,1]$$

We have $g(F,a)$ as the retrieval status value (RSV) for a query q of one term (term t) with query weight a in terms of document d, which has index term weight $F(d,t)$ for the same term t. That is to say, g is the evaluation of document d in terms of its estimated relevance with respect to this one-term query q. Kraft, et al. [Kraft94a] consider a variety of forms for g, striving to allow the function to reflect the semantics of the query. The function g can be interpreted as the evaluation of the document in question along the dimension of the term t if the actual query has more than one term. While many researchers have considered models that do not use Boolean logic for the queries [Salton89], virtually all commercial systems use Boolean logic for multi-term queries. For this case, let

$$e: [0,1]^* \longrightarrow [0,1]$$

be the RSV for a Boolean query of many terms, where each term in the query is evaluated as a single-term query against the document and then the total is evaluated using fuzzy Boolean logic (e.g., AND uses a max function, OR uses a min function, and NOT uses a one-minus function). This notion of allowing e to be a function of the various g values is based on the criterion of separability [Kraft83].

We note that the evaluation functions, g and e, in effect measure the similarity of the document to the query. It has been shown that one can measure the similarity of one document to another via an analogous approach [Salton89; Salton93; Salton94].

Evaluation of a retrieval system is often based on two factors, recall and precision. Recall is the percentage of the relevant documents that were retrieved, while precision is the percentage of the retrieved documents that were relevant. Of course, other factors, such as cost, can enter into the evaluation of a retrieval system. Moreover, it is well known that topicality is but one key facet of relevance; users are influenced by other factors such as language, timeliness, and appearance.

Last Summer's Fuzzy Boolean Retrieval Project

Because last summer's project and this summer's project are so intimately linked, we need to summarize the background, methodology, and results of last summer's work before going into this summer's work. The concept of using keywords to determine a relevant subset of the EDC entries in

response to a user's query was the focus of last summer's project. It made a good deal of sense to allow users to form a query and get a subset of entries to explore. The hypertext navigational aids of CASHE:PVS, perhaps extended by this summer's project, could then allow the user to go from there and find the proper design answers that were being sought. It is the extension of those hypertext navigational aids, based on term frequencies in EDC entries, that is the focus of this year's summer project. In order to understand the reasons for the approaches taken with this project, we now need to consider the environment in which we were working.

The first step was to get a copy of the Design Checklist for use in our retrieval experiments. The file had portions of the topic headers given in the hierarchical classification of the queries, as well as the queries themselves. It was manually edited and then suitably processed to generate the full headers, followed by the queries in the Design Checklist, with links to the appropriate EDC entries. Inconsistencies were noted and fixed.

The next task was to generate an inverted file or index of the keywords in the headers and in the queries under those headers. This meant that a merged file of edited full headers and queries would be needed, with stop words removed. The original list of stop words is given in Appendix II. The canned program we used for this step is from [Frakes92; Fox, C., "Lexical Analysis and Stoplists"]. The keyword list was stemmed, using Paice's stemming mechanism, then sorted in ascending alphabetic order, and duplicates were counted to determine frequency.

We then generated a mechanism for users to enter requests for the appropriate Design Checklist query or queries. Initially, only one term could be requested, but the user could specify whether or not he/she wished the search to use the term as it was or to stem it when making comparisons in the index. Once the system found a match (or possible matches if stemming was used) between the request term and the keywords in the inverted file (index), the system then searched through the list of query numbers and matched them with the query numbers in the query file to produce the query and the EDC number of the entries that answered the query. This was a very slow process but it established the fact that the keyword search could be accomplished successfully, which was important in this experimental procedure. However, efficiency would have to be improved in future generations of CASHE:PVS. A list of the EDC numbers attached to the queries was then provided. Experiments showed that additional keywords in the EDC entries that one might predict would be useful in describing these entries were present.

Weights were then added based on Salton's IDF-based formula in order to improve retrieval performance by representing partial relevance and ranking the output. It was decided to incorporate Salton's IDF-based formula, as presented above, for the weights. The inverted file and query mechanisms were modified to take these weights into account. The query mechanism was then modified to take advantage of the indexing weights. One noted that the listing provided the same results in our experiments in terms of the queries that were found in the non-weighted version. Further, the queries were

not ranked as presented; this feature should be added with the choice of either being in order of query number or of weight being given to the user.

Another experiment showed that searches on terms in relevant EDC entries were useful in searching for other entries. This showed that automatic generation of links between EDC entries was indeed possible and probably desirable. A list of additional terms in the output ranked by frequency showed that term correlations were useful. In addition, the topic headers from the Design Checklist were displayed along with the queries for context. Finally, the Boolean AND and OR connectives were added for simple, two-term queries due to time constraints. Fuzzy logic was employed for the weights [Kraft83; Kraft85; Kraft94a], so two terms were ANDed via the minimum of the weights of each term as assigned to a given Design Checklist query, while the maximum of the weights of each term was used when the terms were ORed in the request. Of course, a term that did not occur at all in the Design Checklist query or appropriate header was assumed to have a weight of zero.

This Summer's Automatic Generation of Hypertext Links Project

Let us suppose that one had found a relevant EDC entry, perhaps by the use of the query processing mechanism outlined above via the Design Checklist. The question becomes one of allowing the user to navigate, i.e., browse, through other EDC entries that have similar content in that they provide additional relevant information to the user. CASHE:PVS has some of those capabilities now. One can use the EDC Table of Contents to locate similar entries. One can use the back-of-the-book index to find other entries of interest if one can specify a keyword of interest.   More to our focus here, one can use the cross references field, if present, to find related entries, which is directly related to hypertext links. Or, one can automatically generate a series of entries through which the user may select to browse.

The real issue becomes one of determining the appropriate linkages between documents. The links currently in place have been put there manually (in the cross references field of the entries). It is not at all clear that these are comprehensive in that other links might be as appropriate if not more so. The question is how to do this automatically in a reasonable manner that reflects true entry similarity based on user preferences and needs.

Salton, et al., [Salton93; Salton94] suggest that the term weights used to index documents for retrieval purposes could be used to measure document similarity for use in automatically generating hypertext links. A document is linked to another document if the similarity between the first and second documents is large enough. While document similarity has been used to cluster documents for retrieval [Salton89], the idea of extending this for hypertext links is new. Salton and his co-authors obtain good results using this method to link articles in an encyclopedia, although they find that performance is improved if one can break down large articles into separate sections, subsections, and even paragraphs or a handful of contiguous sentences.

It was decided to try to apply this approach to the CASHE:PVS system. We had already seen how

a weighted document representation based on keyword frequencies could be employed for retrieval purposes. The issue now was how well document similarities based on those frequencies would parallel proper EDC entry linkages. Our basic unit was the EDC entry.

Data Preparation

First, we felt that the entire EDC would be too large to try as a prototype over the summer. Thus, we decided to narrow our focus to some sections and subsections of the EDC that focused on audio issues. We selected the following sections and subsections:

2. Auditory Acquisition of Information
    2.1 Measurement of Sound
    2.2 Physiology of the Ear
    2.3 Detection
    2.4 Discrimination
    2.5 Temporal Resolution
    2.6 Loudness
    2.7 Pitch
    2.8 Localization
6. Perceptual Organization
    6.4 Auditory Perceptual Organization
8. Human Language Processing
    8.3 Intelligibility of Speech
    8.4 Intelligibility of Altered Speech
10. Effects of Environmental Stressors
    10.3 Noise

The complete list of these sections and subsections, along with the titles of the corresponding entries, is found in Appendix I to this final report. The appendices to this year's final report are not attached to it; but are available from the contact person for the project, whose name and address are given below:

Donald L. Monk, Program Manager
Armstrong Laboratory
Crew Systems Directorate
Human Engineering Division (AL/CFHD)
Wright Patterson AFB, OH 45433-6573
(513) 255-8814
dmonk@falcon.aamrl.wpafb.af.mil

This set of sections, subsections, and entries represented 128 different files , one for each of the 114 individual entries, twelve for each of the separate subsections (2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.08, 6.04, 8.3, 804, and 10.3), one for section 2. (which was the only section used completely in our prototype), and one for the entire prototype. This latter file is treated as a master file because it contains all of the text in the prototype and thus had all of the keywords that would be used. Since each of these files represents a section, the subsections and individual entries were created manually by editing out what was not desired for each file. Each file was transferred for further processing. A list of the names of the files was also stored The file names are structured as follows: file edc02.txt has section 02, file edc0201.txt has section

0201, and file edc020101.txt has entry 020101.

    The next thing to do was to identify all of the individual words and eliminate the stop words. The mechanism used was the same one we used last summer (the canned program written for the chapter on stop lists in [Frakes92; Fox, C., "Lexical Analysis and Stoplists,", Chapter 7]). A copy of the entire list of 423 stop words is in Appendix II. Following this, the list of non-stop words (keywords) for each file was stored in a set of temporary files. For each of these temporary files, the keywords would then be stemmed to eliminate suffixes and find common roots. The stemming procedure that we employed was based on a set of rules to determine how to strip off suffixes as given in the canned program that accompanied [Frakes92; Frakes, W., "Stemming Algorithms," Chapter 8]. A set of temporary files was created to hold the root terms for each keyword file that resulted from the stemming program.

    The next thing was to sort each of these stem files in ascending, alphabetic order. Finally, the inverted file or index, to each file had to be created. The temporary files of sorted keywords were then processed by passing each of them in turn past a computer program we developed to generate a list of each keyword stem and a count of how often it occurred in the file at hand. We note that in total there were |T| = 2,857 keywords used in our prototype.

Term Weights and Similarities

    Now the term weights w_{it} must be calculated. We decided to use a variant of Salton's [Salton89] method, based on the inverted document frequency metric. Thus,

$$w_{it} = f_{i,t} * \{ \log [N / N_t] + 0.1 \}$$

where $f_{i,t}$ is the count for term t occurs in entry index file i, N is the number of entries in the collection (128), and N_t is the number of entry index files in the collection in which term t occurs at least once. The IDF is the log factor in the formula. Note that we did not bother to normalize the weights, since the similarity measure does not require it. Moreover, we decided to add a constant, 0.1, to the IDF so that no weight would be zero. In retrospect, this may have been too large, and calibration is definitely an issue for further study.

    In order to test the system, a subset of the entry files was selected to be run. The EDC entries 8.401, 8.402, 8.403, and 8.404, along with the subsection 8.4, were selected as a sample for testing and debugging. The sample had a total of |T'| = 412 keywords used. The weights for these sample files are given in Appendix III. Next, we needed to calculate the similarity of one document with another. This is analogous to retrieval, where Salton [Salton89] uses his cosine mechanism to calculate the similarity of a document to a query or to calculate the similarity of documents for the purpose of clustering. Following Salton's [Salton93, Salton94] suggestion to use this similarity to link entries on the basis of similarity, we get

$$\text{sim}(i,j) = \text{similarity of entries i and j} = \sum_{t=1}^{|T|} w_{it} * w_{jt} \Big/ \sqrt{\sum_{t=1}^{|T|} w_{it}^2} \sqrt{\sum_{t=1}^{|T|} w_{jt}^2}$$

where |T| = the number of terms. |T| can be found by examining the size of the index file for the EDC prototype file that has the concatenation of all sections, subsections, and entries (for our sample, we look at the size of the index file for the 8.4 subsection file). Thus, for any entry i, we can show the similarity to any other entry j. Note that sim(i,i) = 1.000 by definition (an entry is perfectly similar to itself). Also, the similarity measure is symmetric, so sim(i,j) = sim(j,i) for all entries i and j. This also allows us to generate a ranked list of the entries in order of their similarity to a given entry to produce a list of hypertext links generated automatically. A copy of the similarities for the sample is in Appendix IV. A copy of the ranked list of entries based on similarities in descending order for each entry in the sample is also in Appendix IV.

It should be mentioned that the Smart system [Salton89] from Cornell University, a text retrieval system, was obtained via FTP. A copy of the compressed file is available, but needs to be transported to a Unix-based computer with a C compiler. However, it was decided not to employ Smart for this project, reserving it for possible later use. One can obtain an additional copy of Smart via anonymous ftp to ftp.cs.cornell.edu (login id is anonymous, password is the user's login id and email address), changing directory (cd pub/smart) and getting the desired compressed file (smart.11.0.tar.Z). In addition, this same directory has several test document collections one can also obtain via ftp. The compressed Smart system file must then be ported to a Unix machine and decompressed (zcat smart.11.0.tar.Z I tar xf -). One must then install Smart by making smart/11.0/install the working directory, updating the makefile, and then running make. One can also send a message to smart-people-request@cs.cornell.edu to become a member of the electronic bulletin board for Smart.

Modifications

With these results, an analysis showed that some of the entries in the sample had cross references to other entries in the sample that were not listed as being among the entries seen as most similar to those initial entries. Thus, it was decided to try to experiment a bit with the mechanisms that automatically generated the links. First, we decided to remove the key references and cross references sections from the sample entries and subsection files. Rerunning the stop word, stemming, sorting, and count mechanisms to regenerate the inverted files, then rerunning the generation of the weights and similarities revealed a bit of change in the ranking of the linked entries. This produced a reduced total of 356 keywords. The results (similarities and ranked lists) are in Appendix VI. While there was a small change for the sample data, it is speculated that in general this might not be helpful because terms in the cross references (which give the title of the entry being cross-referenced) were no longer present, which could result in moving these manually-linked entries further apart rather than making them more similar.

Another experiment changed the stop word list, deleting some words that could be important keywords, while adding some words that are not keywords. This moved the stop word list to be more in line with what the CASHE:PVS project uses. The modified list of 374 stop words, with the changes

indicated, is in Appendix VII. This produced a revised |T| = 2,909 keywords in the prototype, keeping the key references and cross references intact. The similarities for the entire subset of entries for our summer prototype are given in Appendix V.

The result of the similarities for the sample using the original text (with the key references and cross references sections left in the files) was changed only somewhat when comparing it to the result with the original stop word list. This is not too surprising given the small sample size of the sample, so that the indexes were not much affected by this experimental change. Thus, the results of the similarities for the sample using the modified text (with the key references and cross references sections omitted in the files) again were nearly the same as the results with the original stop word list, no doubt for the same reason.

The similarities for the section and subsection files for the entire prototype with the new stop word list have been analyzed further by Don Monk. First, Appendix VIII shows a comparison of the EDC cross references (the manual links) and the top ten entries based on the automatically generated similarities. This analysis was conducted for a subset of EDC entries (2.101, 2.102. 2.103. 2.104, 2.105, 2.501, 2.502, 2.503, 2.504, 8.312, 8.313, 8.314, 8.315, 8.316, 8.317, 8.318). We see that in a few cases (2.101, 2.502, 8.313, 8.318) the results are perfect in that the rank order of the similarities matches identically with the cross referenced entries. In only one case (8.315) did the similarities look poor in comparison to the cross references. In most cases, the similarities seemed to perform well, assuming that the cross references were a good basis for evaluation. Appendix VIII also has the similarities for these selected entries.

The issue of treating the subsections and sections as entries is problematic in that they are really collections of entries. However, it was decided to look at how the prototype index file, the section 2. index file, and the twelve subsection files linked together. The result of that analysis is in Appendix IX, where the similarities are given. We can see that there is not as much similarity among the subsections in section 2., but that subsections 8.3 (intelligibility of speech) and 8.4 (intelligibility of altered speech) have a high degree of similarity.

One final experiment was run at the suggestion of the contact person, Don Monk. We isolated the key terms field for each of the 128 entry files. The notion was that the set of key terms in an entry represented a concept, and one should try to find other entries with similar concept descriptions. Blanks and hyphens (dashes) were removed from the key words forming long words from phrases (e.g., "whitenoise"). Then, without using stop words or stemming, we generated a inverted file or index for these keywords from the key terms field. Finally, we generated the similarity matrix and document rankings for this case. The similarities for the subset of entries mentioned above (2.101, 2.102. 2.103. 2.104, 2.105, 2.501, 2.502, 2.503, 2.504, 8.312, 8.313, 8.314, 8.315, 8.316, 8.317, 8.318) are given in Appendix X. In addition, Appendix X shows how the similarities, based on the key term field, ranks in comparison to the manual links in the cross references. It is clear that the key terms in the prototype are

good indicators of concepts and that the generated links generally make sense. A comparison of the rankings based on the key terms versus the entire entries are also seen in Appendix X. The key terms alone do not do as well as the entire entries, which is not surprising. However, there seems to be a relationship between the two different rankings, albeit not a perfect one. It is anticipated that the key terms might work better when the entire EDC is used. One also wonders what other fields might do well. The importance of this idea of key terms as concepts is due to the fact that this field is considerably shorter than the entire entry and the total number of keywords in the filed considerably fewer than for the entire entry set. Thus, one could augment the static (cross references) links and the automatically generated links based on entry similarities, both of which are static in that they are generated once and for all before users query the system. The augmented links would be dynamic, involving user feedback. A user would identify a subset of entries as being relevant among those retrieved or linked to retrieved entries, and the union of the key terms in those entries could be used to find links dynamically to more potentially relevant entries in real-time. This certainly deserves further exploration.

One central theme runs through these results. More research is needed to calibrate and fine tune the process. One issue that may be a fundamental cause of the lack of better results is that the domain is too small. Salton, et al. [Salton93, Salton94] use an encyclopedia as a base, which a tremendously wide variance in subject coverage. On the other hand, CASHE:PVS uses the EDC entries, which are all about human performance studies. Moreover, the prototype was a subset of approximately ten percent of the entries, and all were related to sound; and the sample was a subset of four entries from the prototype, all related to intelligibility of altered speech. With so little difference in content among the different entries, the calculation of the similarities may be overly influenced by noise, i.e., fluctuations in term weights for terms that are not all that important in establishing the presence or absence of a hypertext link.

Summary, Conclusions, and Future Work

It is clear that weighted (fuzzy) keyword retrieval works well in a variety of textual database environments. It is also clear that this approach holds much promise for automatic generation of hypertext links. It has been shown that these possibilities for such a system for CASHE:PVS definitely did, and still do, exist. A system can be constructed that would be able to let users quickly and easily find and use relevant entries, some found via retrieval and others via links to the retrieved entries.

However, a caveat has to be expressed. The retrieval programs developed here were done to quickly demonstrate the feasibility of such a system. The step-by-step development outlined above in a time sequence needed to be submitted to a much more rigorous software engineering approach. A much more integrated system should be developed with more general capabilities along with appropriate systems and user documentation. For example, the request processing mechanism needed to be able to handle requests of arbitrary complexity, including negation (NOT) and as many terms as the user desires. In addition, the user should be able to refine a request, adding terms to add queries to or to delete queries

from a given search output. The user should be able to rank the output on a variety of criteria (e.g., weights, EDC entry numbers). In addition, the remaining EDC entries need to be added to the hypertext link generation mechanism.

Moreover, as Don Monk has suggested, more is needed than the static hypertext links generated either manually in the cross references field or automatically as discussed above. Dynamic links can be formed based on the concept of relevance feedback [Salton89] where the system attempts to provide additional entries based on the user noting which among the entries viewed by the user were good and which were not. Suppose a user has identified a small subset of entries that are of interest, perhaps using the retrieval and link mechanisms discussed above. One could use the keywords, perhaps in the key terms field, found in those entries as a conglomerate to represent a concept on the fly. One could then calculate similarity weights for the other entries and look at the ranked list to see which new entries are worth viewing given the user's needs as expressed in the entries seen and found to be relevant. One wonders if one could add a negative factor if keywords occurred in entries found to be non relevant. Moreover, one wonders if the keywords should come from just the key terms field, or if title, general description, table of contents heading, or other fields could be of importance.

Obviously, more research is needed. First, for retrieval there are a variety of access points available to the CASHE:PVS user. One could use the Design Checklist header outline, the Design Checklist queries, the EDC Table of Contents outline, the EDC back-of-the-book index, and the impending keyword index of the EDC entries that have been developed. One needs to correlate these to see which ones are most effective and how they could be integrated into a whole system for easy and efficient access for users.

Second, to take fuller advantage of the notion of document components, which should be done to become more in line with the successful work that Salton [Salton93, Salton94] has done, we need to explore weightings for terms depending on where the terms were found, i.e., in which of the different components (e.g., title, table of contents heading, figure or illustration caption, the various fields such as constraints or experimental validation) the term was found. Thus, instead of saying that the term "ear" was found six times in a given entry, we might note that it was found once in the key terms field, once in the title field, twice in the general description field once in the cross references field, and once in the key references field). If we weight these fields as 1.00 for key term, 0.95 for title, 0.85 for general description, 0.05 for cross references, and 0.01 for key references, we get a score of 3.71.

The issue of calibration is an important one. The determination of optimal parameters and mechanisms so performance of the retrieval and hypertext link generation processes is maximized is obviously critical. Possible weights for the different fields in which keywords occur is one place for calibration. Another is the generation of the weights themselves; we noted before that we added a constant of 0.1 to the IDF factor; calibration must be done to see if that factor is desirable and if so what should its

value be.

Fourth, the issue of which terms to put in the stop word list was raised this summer, and there are no real answers in general for what is optimal. Further explorations of term frequencies in the CASHE:PVS system to help determine the stop words list might prove fruitful.

A fifth area is the issue of the size of the hypertext link generation problem. For example, the size of the prototype, which is but a subset of the entire EDC, makes display of the entire ranked list of entries too large. For our prototype, we have 128 entry files (treating each section and subsection as an entry, as noted before). We need to determine a mechanism for determining a cutoff, either in terms of the number of entries (displaying only the top K entries linked to the given entry) or in terms of the similarity (displaying only the entries linked with a similarity to the given entry of no less than α). The modifications to the computer program that calculates and displays the weights and similarities would be trivial to make. A key issue is how to determine the best value of K and/or α, i.e., another calibration problem.

Finally, one needs to confront the entire CASHE:PVS system with real end-users, i.e., the designers for whom it was intended. For example, there is the concern of how to evaluate the performance of the retrieval and hypertext link generation mechanisms. This is especially true of the hypertext links. Does one simply compare the similarities to the suggested cross references? If so, are additional automatically generated links worthwhile? Yet, the mechanism should certainly indicate a strong similarity of entries that have a cross reference relationship. It is clear that the system needs to be confronted with users to test these links.

Acknowledgments

AFB, and Mead Data Central to informally discuss the fuzzy retrieval and hypertext project for CASHE:PVS.

Finally, work by colleagues, including the software accompanying [Frakes92] and the software for stemming developed by Paice, must be properly acknowledged.

References
[Boff86]
     Boff, K. R., L. Kaufman, and J. P. Thomas, *Handbook of Perception and Human Performance*, vols. I and II, New York, NY: John Wiley & Sons, Inc., 1986
[Boff88a]
     Boff, K. R. and J. E. Lincoln (eds.), *Engineering Data Compendium: Human Perception and Performance*, vols. I, II, and III, Wright-Patterson Air Force Base, OH: Human Engineering Division, Harry G. Armstrong Medical Research Laboratory, 1988
[Boff88b]
     Boff, K. R. and J. E. Lincoln (eds.), *User's Guide: Engineering Data Compendium, Human Perception and Performance*, Wright-Patterson Air Force Base, OH: Human Engineering Division, Harry G. Armstrong Medical Research Laboratory, 1988
[Boff91a]
     Boff, K. R., D. L. Monk, and W. J. Cody, "Computer Aided Systems Human Engineering: A Hypermedia Tool," presented at SOAR '91, Space Operation Applications and Research, NASA-Houston, July, 1991
[Boff91b]
     Boff, K. R., D. L. Monk, S. J. Swierenga, C. E. Brown, and W. J. Cody, "Computer-Aided Human Factors for Systems Designers," presented at Human Factors Society annual meeting, San Francisco, CA, September, 1991
[Cona93]
     Cona, T. R. and D. L. Monk, "Bringing Human Performance Data to the Design Table," to be presented at Human Factors and Ergonomics Society annual meeting, Seattle, WA, October, 1993
[Frakes92]
     Frakes, W. B. and R. Baeza-Yates (eds.), *Information Retrieval: Data Structures and Algorithms*, Englewood Cliffs, NJ: Prentice Hall, 1992
[Klir88]
     Klir, G. J. and T. A. Folger, *Fuzzy Sets, Uncertainty, and Information*, Englewood Cliffs, NJ: Prentice Hall, 1988
[Kraft93]
     Kraft, D. H., "An Exploratory Study of Weighted Fuzzy Keyword Retrieval for the CASHE:PVS System," Final Report for Summer Faculty Research Program, Department of Computer Science, Louisiana State University, Baton Rouge, LA, 1993
[Kraft85]
     Kraft, D. H., "Advances in Information Retrieval: Where is That /#*%@^ Record?," in Yovits, M. (ed.), *Advances in Computers*, v. 24, Academic Press, New York, 1985, pp. 277-310
[Kraft94a]
     Kraft, D. H., G. Bordogna, and G. Pasi, "An Extended Fuzzy Linguistic Approach to Generalize Boolean Information Retrieval," accepted by *Journal of Information Sciences*, 1994
[Kraft91]
     Kraft, D. H. and Boyce, B. R., "Approaches to Intelligent Information Retrieval," Technical Report 91-020, Department of Computer Science, Louisiana State University, Baton Rouge, LA, 1991

[Kraft83]

Kraft, D. H. and D. A. Buell, "Fuzzy Sets and Generalized Boolean Retrieval Systems," *International Journal of Man-Machine Studies*, v. 19, 1983, pp. 45-56; reprinted in Dubois, D., H. Prade, and R. Yager (eds.), *Readings in Fuzzy Sets for Intelligent Systems*, San Mateo, CA: Morgan Kaufmann Publishers, 1993

[Kraft94b]

Kraft, D. H., Petry, F. E., Buckles, B. P., and Sadasivan, T., "The Use of Genetic Programming to Build Queries for Information Retrieval," IEEE Symposium on Evolutionary Computation, Orlando, FL, 1994

[Monk92]

Monk, D. L., S. J. Swierenga, and J. E. Lincoln, "Developing Behavioral Phenomena Test Benches," presented at Human Factors Society annual meeting, Atlanta, GA, October, 1992

[Miyamoto90]

Miyamoto, S., *Fuzzy Sets in Information Retrieval and Cluster Analysis*, Boston, MA: Kluwer Academic Publishers, 1990

[Petry93]

Petry, F. E., Buckles, B. P., and Kraft, D. H., "Optimizing Boolean Queries with Relevance," presented at the American Society for Information Science (ASIS) annual meeting, *Proceedings*, Columbus, OH, October, 1993

[Salton89]

Salton, G., *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Reading, MA: Addison-Wesley Publishing Co., 1989

[Salton93]

Salton, G. and J. Allan, " Selective Text Utilization and Text Traversal," Hypertext '93, Seattle, WA, November, 1993, *Proceedings*, New York, NY: ACM, 1993

[Salton94]

Salton, G., J. Allan, C. Buckley, and A. Singhal, "Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts," *Science*, v. 264, June 3, 1994, pp. 1421-1426

Apoptosis, Advanced Glycosylated End Products, Nitric Oxide as related to Hyperbaric Oxygenation Therapy

B. DeSales Lawless

Professor

Department of Science and Mathematics

Fordham University College at Lincoln Center

New York, N.Y., 10023

Final Report for:

Summer Faculty Research Program

Armstrong Laboratory

Apoptosis, Advanced Glycosylated End Products, Autofluorescence
and Nitric Oxide as related to Hyperbaric Oxygenation Therapy

B. DeSales Lawless

Professor

Department of Science and Mathematics

Fordham University College at Lincoln Center

New York, N.Y., 10023

Abstract

We studied the free radical Nitric Oxide and its effects on apoptosis (programmed cell death) and non- enzymatic glycosylated end products (AGEs). The work involved the antigen-presenting dendritic cell, cells from an environment of high glucose concentration, and rat cells under the stress of microwave radiation. The Griess Reagents were used in the assays of nitric oxide. Flow cytometry was used in much of this work.

Apoptosis, Nitric Oxide and Advanced Glycosylated End Products as related to Hyperbaric Oxygen Therapy

B. DeSales Lawless

Introduction

Nitric Oxide has become the focus of much research. This short-lived free radical functions as a neurotransmitter, is involved in vasodilation, inflammatory responsis, autoimmunity, platelet aggregation, etc. High levels of nitric oxide are associated with hypotension, endotoxic shock, and in the formation of carcinogenic N-nitrosamines. Its cytotoxicity role in tumor and pathogen distruction is well documented. Evidence suggests that it is responsible in some instances for apoptosis (programmed cell death.)

Nitric oxide is produced from L-arginine with the enzyme Nitric Oxide Synthase (NOS, EC 1.14.13.39). This enzyme has a constitutive isoform that has been isolated from bovine endothelium, There is also an inducible isoform found in mouse macrophages. Both forms require NADPH and oxygen. Plants produce No using Nitrate Reductase (NR).

One segment of this study is to determine is i-NOS is functional in human dendritic cells. Another phase is to determine its role in apoptosis.

Methodology

<u>Nitric Oxide Assay</u>.   Nitric oxide content in solutions is determined with the Griess Reagents (0.8% sulfanilic acid and 0.5% N,N-dimethyl-alpha-naphthylamine, both in 5N acetic acid). Baxter Medical Corp., Sacramento, CA. The assay colorimetrically measures nitrite concentration which correlates with NO.

In a typical experiment, 10,000 cells/100ul are placed in a well of a 96 well microtiter plate with 100 ul of each Griess reagent. The plate is incubated 10 min at 37°C. Absorbance is read at 546 nm on a BioRad model 450 microplate reader.

<u>Flow Cytometry</u>.   Flow cytometry analyses for apoptosis and phenotyping were performed with FACScan Flow Cytometer (Becton-Dickinson, San Jose, CA.) Cells were excited with a single 488 nm argon laser and 5,000-10,000 events were collected. Data were analyzed with the Becton-Dickinson Lysis Software.

For apoptosis studies cells were fixed for twenty-four hours in 70/5 ethyl alcohol at 4°C according to the procedureof Coligan, et al in <u>Current Protocols in Immunology.</u> The DNA of the cells was stained with 1 ml solution of propidium iodide (50 ug/ml) containing RNase-A (100 U/ml). The instrument was calibratead using chicken erythrocytes and calf thymus cells which contained 33% and 80% of the 2N DNA of human cells respectively.

24-4

Data graphics were presented with dot plots and histograms including appropriate statistics. Autofluorescence was estimated using dot plots with forward scatter vs side scatter. Cells identified as apoptotic were identified by hypodiploid fluorescence.

Phenotyping for CD4+(T-helper)/CD8+T-suppressor) cells was effected with the Becton-Dickinson Simultest fluorescent antibody mixture. Anti-CD4+ was fluoroisothiocyanate conjugated; anti-CD8+ was phycoerythrin conjugated. Cells were incubated on ice for 30 min with 1 ul of the fluorescent antibody mixture. Statistics were obtained with quadrant analysis of dot plots. Phenotypic analysis of other surface markers was made in a similar way using the appropriate fluorescent-conjugated primary antibody.

Dendritic Cell Preparation. Dendritic cells are prepared by the procedure of Bhardwaj, et al. Granulocytes and erythrocytes are removed from heparinized blood by Ficoll-gypaque centrifugation. T-cells are then removed from the mononuclear cells by sheep red blood cell rosetting. The ER⁻ fraction are kept in plastic dishes for 60 min and nonadherent cells are removed. The non-adherent cells are depleted of monocytes by panning on Ig-coated plates and layered on 14.5% metrizamide. B cells and NK cells pellet and dendritic cells float. The fluorescence activated cell sorter is used to negatively remove not dendritics; positive markers yield a dendritic fraction >95% pure.

24-5

Investigations

Dr. Stephen Peiper of the Brown Cancer Center, University of Louisville, Kentucky, collaborated in our work. He prepared dendritic cells from his own blood. He was given the protocol used at Rockefeller University and cited the Bhardwaj paper. He planned to send the cells overnight to our laboratory. Dr. Peiper's cells did not arrive in time for us to complete our plans; he will send the cells instead to our laboratory in New York. We received prepared sheep red blood cells (neuraminidase treated) and metrizamide from Rockefeller University so that we could make our own preparation of dendritic cells there. Circumstances prevented us from isolating the cells there.

The work with dendritics had the highest priority for our summer investigations. We were to determine is the i-NOS gene is operable or if we could successfully induce it. Our premise was that induced nitric oxide would fragment any virus carried on these antigen presenting cells before the virus could be delivered to cells having appropriate antigen receptors.

We will attempt to shorten the dendritic cell preparation by using rat spleens, removing as many cells as possible by mincing and washings. Then, following the procedure of Crowley, et al. use collagenase to obtain the connective tissue sequestered dendritics.

Microwave radiation stress. Eight pairs of male Sprague-Dawley rats were used to study the effects of microwave stress radiation on serum nitric oxide content and PBMC response to lipopoly-saccharide stimulation. Individual animals were exposed to 35-GHz continuous wave RFR at an incident power density that yielded a whole-body average specific absorption rate of 13W/kg in the E orientation.

Results: Serum Nitrite (NO) mg/L):

Treatment group	average	standard deviation
microwave	6.06	3.80
control	11.14	5.64

There are less than five chances in a thousand that no difference exists between the two treatment groups.

PBMC measurements. Nitrites (mg/L)

Treatment group	pre LPS		post LPS		% Increase	
	avg	S.D.	avg	S.D.	avg	S.D.
microwave	10.05	8.35	19.69	14.51	123	122
conrol	3.78	2.61	3.94	2.10	22	45

Thare less than five chances in a thousand that no difference exists between the two treatment groups for pre LPS, post LPS and % increase values.

Autofluorescence. Peripheral blood mononuclear cells were cultured in RPMI medium containing antibiotics, and 10% fetal calf serum. A similar population was cultured in the same medium to which 300 mg% glucose was added. The cells were viewed on FACScan using dot plots and forward scatter vs side scatter.

24-7

When macrophage population was gated there was detectable autofluorescence from Fl-1, Fl-2, Fl-3. There was no fluorescence from the lymphocyte population. This was as expected.

There was no detectable autofluorescence from lymphocytes cultured in high glucose concentration. There was no increase in autofluorescence in the macrophage population in high glucose.

Conclusion: Non-enzymatic advanced glycosylated end products are formed only after an exposure lasting more than overnight. We learned that exogenous AGE compounds can be formed with 20-60 day incubation of protein and glucose.

Nitric Oxide Production in Microgravity. Dr. Marian L. Lewis of NASA at Huntsville sent us samples of sera from human PBMC which had flown on the Shuttle S-60 space flight. We assayed the sera for NO content and found significant elevation of NO in the sera of cells flown for eight days on the space flight.

Results: Data is given in optical density readings using the Griess reagents. In the standard curve using sodium nitrate and wave length 570 nm: o.d. 1.0 = 40 ug/ml.

Ground control cells	Flight experimental cells
0.267	0.424
0.256	0.442
0.254	0.512
0.052	0.308
0.051	0.224
0.043	0.143
0.036	0.217

<u>Diabetes mellitus</u>.   Remnant blood is obtained following routine
blood tests for diabetic and non-diabetic patients undergoing
hyperbaric oxygenation therapy. These patients undergo hyperbaric
oxygenation therapy for problem wounds.   PBMCs are isolated by
Ficoll-hypaque density gradient cengtrifugation.   Cells are
analyzed on FACScan (Becton-Dickinson) utilizing an Argon-Ion laser
emitting 488 (300mW). Cell shape and size are analyzed with light
forward and right angle side scatter. Fluorescence is measured at
530 nm, 585, nm and 650 nm. Phycoerythrin and fluoroisothiocyanate
fluorescently labelled antibodies are used for phenotypic
analyses.

Apoptosis is identified using FACScan using propidium iodide
to stain chromosomal DNA.   Hypodiploidy determines apoptosis
(Schmid, 1994).

Activated T-cells are identified by anti-HLA-DR/CD3 and
Interleukin-2 receptor (Tac) markers. Functionality is checked by
mitogen stimulation followed by 3-H thymidine incorporation assay
and cell scintillation counter measurements for proliferation
(Bhardwaj, 1992). Increases in T-cell numbers relative to controls
indicate T-cell activated responses.

CD11b and CD 18 fluorescently labelled antibodies are used to
measure changes in cell adhestion protein expression. (Michell,
1980).

Procedure: We received the bandages which wrapped the non-healing wounds of the persons involved in this study. Cells were washed from the bandages and resuspended in PBS (phosphate buffered saline). Cells from diabetic blood was compared with cells from non-diabetics. The blood samples were first analyzed for glucose concentration.

Results of these studies are are now being analyzed and will be reported separately.

Summer Research at Brooks Air Force Base...A Summary:

Upon arrival at Armstrong Laboratories we were advised that the building assigned to our work was not available because of an asbestos removal problem. Major Kemper of the Air Force provided us space in the Hyperbaric Oxygen Unit. Dr. Edward Piepmeier who joined in our research at Brooks the previous summer agreed to collaborate again. Dr. Johnathan Kiel arranged for us to use the facilities in another building where the flow cytometry instruments had been moved. As the summer progressed, we used some instrumentation at the Air Force Wilford Hall Medical Center. We also joined Dr. Kathy Ryan in the microwave housing.

The Original Proposed Research:

(a) Dendritic Cell Study. Is there an functional gene for Nitrogen Oxide in these human antigen-presenting cells. We would use the Griess Reagents for our assay. We had available the NOS inhibitor, N-methyl Arginine, and some antioxidants: glutathione, NAC, etc.

(b) Autofluorescence. The non-enzymatic advanced glycosylated end products (AGE) found on long-lived proteins and lipids in patients with elevated glucose serum concentrations are known to autofluoresce. We would see if this characteristic could be used in a diagnostic assay or analytically in fractionation. We would see if amino-guanidine and related compounds could lower concentrations of AGE.

(c) Apoptosis. Cells have "normal" life spans (Hayflick Principle). We would look at various cells to induce apoptosis (programmed cell death) and to see if we could delay it. Apoptosis is evidence by observing hypodiploid DNA strands stained with propidium iodide. We would like to keep human T-helper cells alive longer than their normal life span or prevent their undergoing apoptosis when HIV infected. We would attempt to transfect T-helper cells with Bcl-2, a proto-oncogene known to suppress apoptosis. Perhaps we could grow these cells in antioxidant thiol compounds and also obtain these ends. Monomethyl arginine is a competitive inhibitor of NO synthase which also inhibits apoptosis.

(d) Enconucleases. One characteristic of programmed cell death is the activation of an endonuclease which gives rise to DNA fragments of 180-200 base pairs or multiples of these. We would use affinity chromatography by coating microtiter plate wells with various lectins. We would then add serum or supernatant fluids of cells known to undergo apoptosis. If the responsible endonuclease is a glycoprotein, we could begin its characterization.

(e) Will exogenous NO induce apoptosis (or necrosis) in vitro in PBMCs? We can gate on specific cells within the PBMCs after staining with propidium iodide.

(f) We would compare cells grown in hyperbaric oxygen with the results of cells which we have flown in space. We would compare the responses to mitogenic stimulation in hyperbaric oxygen with the results we obtained in microgravity.

New Questions: Among the questions that arose as our research progressed were the following:

(a) We should seek probes for NOS to establish if Dendritic Cells produce Nitric Oxide.

(b) We should prepare exogenous AGE compounds and look for receptors to these compounds on various cell populations. This could be done by fluorescently tagging the AGE molecules.

(c) It is known that a mitogenic autocrine is produced by cells submitted to stress (Dr. J.Kiel). This compound should be identified.

(d)   Prepare Dendritic Cells from mouse spleens by using collagenase to treat the connective splenic tissue after other cells have been depleted. This could shorten the procedural time. Investigations prompted by summer research:n each case.

(e) Prepare dendritic cells and send them to Dr. Lewis. She will grow them in a microgravity environment to determine if it would be worthwhile to add them to Shuttle Experiments.

(f)   Definitely prepare media with various antioxidants (NAC, glutathione, BME, etc. Can such supplements prevent apoptosis. Since exogenous NO will enduce apoptosis, can it be neutralized by such antioxidants.

(g) Compare macrophages of elderly persons with younger persons to estimate the amount of AGEs.

(h)   Answer the question: Why does hyperbaric oxygenation help wound healing in diabetics.

(i) Can AGE be detected on the surface of red blood cell ghosts.

Key Words from the Summer Research:

Hyperbaric Oxygenation Treatment (HBO). Patients are successfully being treated for non-healing wounds with this therapy. Hyperbaric oxygen teeatments are performed at 2.0 to 2.5 atmospheres absolute with a range of 90 to 120 minutes of 100% oxygen breathing. This is at Brooks Air Force Base.

Diabetic Foot Wounds. The most morbid complication of diabetes is foot infection. Diabetic patients account for 50-70% of all nontraumatic amputations in this country and their hospitalization averages 4 to 12 weeks. Diabetes is known to alter function of leukocytes in such a way that achemotaxis, phagocytosis, and intracellular bacterial killing are all diminished. Recent research demonstrates that this lack of function is due in part to increased adhesion to damaged tissue. (Piepmeier, 1993)

Non-enzymatic Advanced Glycosylated End Products. In chronic diabetes excessive accumulation of reactive glucose-protein or glucose-lipid adducts known as AGE explain much of the symptomology of diabetes and aging. Glycosylation of long-lived macromolecules of connective tissue (collagen, fibrin, elastin) is the usual result of excessive glucose concentration in serum. Aminoguanidine 25-15is one inhibitor of this unwanted glycosylation. We would seek to determine if the human immune system responds to AGE by producing anti-AGE antibodies. If they are detected, perhaps the problem could be approached as one of an antibody-complex problem. We would look for exogenous AGE blood samples of various age groups.

The Future:

The many questions that have arisen during this research makes it most worthwhile. There are mountains for us to climb.

REFERENCES

Bhardwaj N, Friedman SM, Cole BC Nisanian AJ. Dendritic cells are potent antigen-presenting cells for microbial superantigens, J Exp.Med, 175: 267-273, 1992.

Corbett TJ, Mehrhof AI, Cohen IK. Possible Therapeutic interventions in insulin dependent diabetes mellitus, Diabetes, 41(8): 897-903, 1992.

Dent GA, Leglise MC, Pryzwansk KB, Ross DW. Simultaneous paired analysis by flow cytometry of surface markers, cytoplasmic or oncogene expression with DNA content, Cytometry, 10: 192-198, 1989.

Freudenthal PS and Steinman RM. The distinct surface of human blood dendritic cells, as observed after improved isolation. Proc Natl Acad Sci USA 1990, 87: 7698-7702.

Hardy RR. Purification and coupling of fluorescent proteins for the use in flow cytometry. Handbook of Experimental Immunology, Blackwell Scientific Publications, Palo Alto, CA, 4th ed, 31.1-31.13, 1986.

Havenith CEG, Breedjik AJ, vanMiert PP, Blijleven N, Calame W, Beelen 'RH, Hoefsmit EC. Separation of alveolar macrophages and dendritic cells via autofluorescence; phenotypical and functional characterization, J Leukocyte Biol, 53: 504-510,1993.

Imani F, Horii Y, Suthanthiran M, Skolnik EU, Makita Z, Sharma V, Sehajpal P, Vlassara H. Advanced glycosylation end product-specific receptors on human and rat T-lymphocytes mediate synthesis of interferon gamma: role in tissue remodeling. J Exp Med, 178: 2165-2172, 1993.

alKassab AS, Raziuddin S. Abnormalities of CD4 T lymphocyte proliferative responses from patients with insulin dependent diabetes. Diabetes Res Clin Prac,10(1): 45-49, 1990.

Krammer PH, Behrmann I, Daniel P, Dhein J, Debatin KM. Regulation of apoptosis in the immune system, Current Opinions in Immunology, 6: 279-289, 1994.

24-15

Misko TP, Moore WM, .Kasten TP, Nickols GA, Corbgett JA, Tilton RG, McDaniel ML, Williamson JR, Currie MG. Selective inhibition of the inducible nitric oxide synthase by aminoguanidine. Eur J Phar, 233(1): 119-125, 1993.

Park E, Quinn MR, Wreight CE, Schuller-Levis G: Taurine chloramine inhibits the synthesis of nitric oxide and the release of tumor necrosis factor in activated RAW 264.7 cells. J Leukocyte Biol, 54: 119-124, 1993.

Piepmeier EH, Todaro R, Kemper GB, Lawless BD. Hyperbaric oxygenation alters leukocyte function. FASEB J. 8 (5): A682, 1994.

Schmid I, Uittenbogaart CH, Keld B, Giorgi JV. A rapid method for measuring apoptosis and dual-color immunofluorescence by single laser flow cytometry. J Immunol Methods 170: 145-157, 1994.

Schmidt AM, Yan SD, Brett J, Mora R, Nowygrod R, Stern D. Regulation of human mononuclear phagocytes migration by cell surface binding proteins for advanced glycation end products. J Clin Invest 91(5): 2155-2168, 1993.

Singer DE, Coley CM, Samet JH, Nathan DM. Tests for glycemia in diabetes mellitus, Annals of Internal Medicine, 110(2): 125-137, 1989.

Stroman R, Piepmeier EH, Todaro R, Kemper GB, Lawless BD, Ryan K, Frei M. Hyperbaric oxygenation alters shock response in rats exposed to 35 GHz microwave irradiation. FASEB J., 8 (5): A845, 1994.

Vlassara H, Fuh H, Makita Z, Krungkrai S, Cerami A, Bucala R. Exogenous advanced glycosylation end products induce complex vascular dysfunction in normal animals: A model for diabetic aging complications. Proc Natl Acad Sci, 89: 12043-12047, 1992.

Wofsy L, Henry C, Kimura J, North J. Modification and use of antibodies to label cell surface antigens. Selected Methods in Immunology, BB Mishell and SM Shiigi, (eds.), W.H. Freeman and Co., 187-304, 1980.

Wu, JT. Review of diabetes: identification of markers for early detection, glycemic control and monitoring clinical complications. J. Clin Lab Anal, 7: 293-300, 1993.

A STATISTICAL METHOD FOR TESTING COMPLIANCE

Tze-San Lee
Associate Professor
Department of Mathematics

Western Illinois University
900 West Adams Street
Macomb, IL 61455

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

August 1994

A STATISTISCAL METHOD FOR TESTING COMPLIANCE

Tze-San Lee
Associate Professor
Department of Mathematics
Western Illinois University

Abstract

In recognition of exposure variability a new statistical method, based on a hypothesis testing approach, is proposed for testing whether it is still in-compliance with the OSHA's standard if the Worker's overall-average exposure level is several times the OSHA's PEL. The asbestos exposure data collected at Lackland AFB is used as an illustration. This data set, consisting of personal samples from six workers exposed to asbestos, possesses several interesting features. Fisrt, the mean exposure level for some workers has changed during the entire exposure period. To detect changes in the mean exposure level a rank cusum chart was applied. Second, the sampling errors between workdays for all workers are highly autocorrelated. An autoregressive transformation was needed to remove the autocorrelated errors from the data. It was noticed that if the autocorrelated errors were not removed, a straight calculation of the exposure variability could be erroneous. Third, it was shown that even if the Worker's overall-average exposure level is several times the OSHA's PEL, it could still be in-compliance with the OSHA's standard. Finally, through the use of the ANOVA model, it was shown that the longitudinal exposure variability is the dominating one which could be from 2 to 8 times the cross-sectional variability. However, the longitudinal variability might be artificially inflated because of a small sample size for each workday. To settle for a more definite conclusion about the relationship between the longitudinal and the cross-sectional variability, more samples need to be collected from more workers.

A STATISTICAL METHOD FOR TESTING COMPLIANCE

Tze-San Lee
Dept. of Mathematics
Western Illinois University

Introduction

Workplace exposures to airborne chemicals are regulated in the U.S. by the Occupational Safety and Health Administration (OSHA) via the promulgation of the permissible exposure limits. However, scrutiny of the OSHA's standards concerning chemical exposures suggests an unfortunate contradiction between the criteria used to promulgate a permissible exposure limit (PEL) and the enforcement of the PEL within the workplace. These limits, usually defined as 8-hour time-weighted average, are enforced as concentrations never to be exceeded Such enforcement procedures lack scientific rigor by defining compliance of employees' chemical exposures in absolute terms. Let us not forget that the PEL was intended to represent the constant level to which a worker is allowed to be exposed over his/her working lifetime. However, workers' exposures are not constant from day to day. Instead, the exposure level is a random variable which follows some kind of probability distribution. Furthermore, OSHA's interpretation of the PEL as an absolute limit will discourage employers from monitoring exposures in the workplace, since the probability of compliance can be maximized by minimizing the number of samples to be collected from the workers.[9]

Any rational enforcing rule has to be statistically based. In fact, OSHA already used a statistical criterion, the so called confidence interval method, in decision making. It issues a citation only after the lower confidence limit (LCL) of a given measurement exceeds the PEL.[8] However, note that this is not a LCL about the true mean exposure level since it only superimposes the measurement error on the value of a single measured exposure. As was pointed out in Nicas,[7] the measurement error is accounted for no more than 13% of the total sampling errors which are the sum of variability of cross-sectional, longitudinal and analytical measurement. When the sample size of the collected data is small, the measurement error is practically negligible compared to other sources of variability in the data.

It is the aim of this paper to propose a new statistical method based on a hypothesis testing approach to test compliance of the worker's average exposure level to the OSHA's standard by taking the exposure variability into consideration. This new method is more rigorous and easy to use than any of the currently available methods. An example is given by using the data collected at Lackland AFB from a team of workers exposed to asbestos. The data set of asbestos exposure concentration possesses some interesting features including there are changes in the mean exposure concentration for some workers throughout the exposure period, and the sampling errors are highly autocorrelated longitudinally. In addition, the cross-sectional (or between-worker) and longitudinal (or day-to-day) variability are calculated by using the analysis-of-variance (ANOVA) model.

Hypothesis Testing Approach

Assume that the random variable, X_c, of chemical airborne concentration is lognormally distributed with mean μ_c (or with geometric mean μ_g) and standard deviation σ_c (or with geometric standard deviation σ_g). The lognormality is a reasonable assumption as long as the joint effects of environmental factors is multiplicative.[3] As was advocated by Rappaport,[10] a relevant index for measuring the hazard effect of chronic exposure to environmental pollutants is the mean exposure level μ_c, not the geometric mean μ_g. To determine whether the worker's mean exposure level is below the PEL of the OSHA standard, a hypothesis testing approach is adopted. It is a common practice in the statistical hypothesis testing to give the benefit of doubt by assuming that the mean exposure level in the

workplace meets the OSHA's standard unless proved otherwise. Thus, both the null and alternative hypotheses are, respectively, defined as follows:

$$H_0: \mu_c \leq PEL$$

versus

$$H_a: \mu_c > PEL \tag{1}$$

To take advantage of the logarithm of the exposure concentration is normally distributed, A more convenient test would involve the parameter μ_ℓ which is the mean of the (natural) logarithmic transformed random variable $X_\ell = \ln(X_c)$. Note that there is a relationship between μ_c and μ_g $(= \exp(\mu_\ell))$ given by[6]

$$\mu_c = \mu_\ell \exp(\sigma_\ell^2 / 2) \tag{2}$$

By using (2), the hypotheses of (1) can be rewritten as follows:

$$H_0: \mu_\ell \leq PEL_\ell$$

versus

$$H_a: \mu_\ell > PEL_\ell \tag{3}$$

where PEL_ℓ is obtained from (2) with $\mu_c = PEL$ and given by

$$PEL_\ell = \ln(PEL) - \sigma_\ell^2 / 2 \approx \ln(PEL) - s_\ell^2 / 2. \tag{4}$$

Now, as usual, the best point estimator for μ_ℓ is the sample mean, $\hat{\mu}_\ell = \bar{x}_\ell$, where

$\bar{x}_\ell = \sum_{i=1}^{n} x_{\ell,i}$, and $\left\{ x_{\ell,i} \right\}_{i=1}^{n}$ are n independent observations of the random variable X_ℓ . Often, the

sampling errors in $\left\{ x_{\ell,i} \right\}_{i=1}^{n}$ are not independent , but instead autocorrelated because $\left\{ x_{\ell,i} \right\}_{i=1}^{n}$ are usually collected during a brief campaign over several days, the estimator $\bar{x}_\ell$ is still unbiased. However, the standard error of $\bar{x}_\ell$ will be affected, and usually underestimated by using the conventional formula. Hence, this feature of having autocorrelations needs to be removed from the data. Here, for simplicity, the autocorrelated errors of $\left\{ x_{\ell,i} \right\}_{i=1}^{n}$ is assumed to follow the first-order autoregressive process, AR(1), with the autocorrelation coefficient ρ. For many practical applications, the autocorrelated errors can indeed approximately be modeled by AR(1).[1] Now, the appropriate formula for the standard error of $\bar{x}_\ell$, denoted by $s_{\bar{x}_\ell}$, is then given by[4]

$$s_{\bar{x}_\ell} = \sqrt{ \left(s_\ell^2 \Big/ n \right) \left[1 + \left(2 \Big/ n \right) \left[\sum_{h=1}^{n-1} (n-h)\rho^h \right] \right] }. \tag{5}$$

An appropriate testing statistics for (3) is defined by

$$t = (\bar{x}_\ell - PEL_\ell)/s_{\bar{x}_\ell}. \qquad\qquad (6)$$

where $s_{\bar{x}_\ell}$ is given by (5). Note that the sampling distribution of t can be approximated by the student t-distribution with n - 1 degrees of freedom. To compute s_ℓ used in (5) from the autocorrelated observations, $\{x_{\ell,i}\}_{i=1}^n$, we should first apply the autoregressive (or Cochran-Orcutt[2]) transformation given by

$$z_i = x_{\ell,i} - \rho x_{\ell,i-1}, \text{ I} = 2,...,\text{ n.} \qquad\qquad (7)$$

to $\{x_{\ell,i}\}_{i=1}^n$ to remove the autocorrelated errors in the collected data. Consequently, s_ℓ is calculated by using the formula given as follows:

$$s_\ell = \sqrt{\sum_{i=2}^n (z_i - \bar{z})^2 / (n-2)}, \qquad\qquad (8)$$

where $\bar{z}$ is the sample mean of $\{z_i\}_{i=2}^n$. Incidentally, the assumption of lognormality on the exposure concentration level can be examined empirically by applying the normal probability plot to the data set of $\{z_i\}_{i=2}^n$. Also, note that since one observation was lost in (7) for removing the autocorrelated errors, the correct degrees of freedom for (6) should modify to be n - 2.

Materials

The data collected at the Lackland AFB from workers exposed to asbestos is used as an illustration how to use the proposed method to test if the worker's exposure level is in compliance with the OSHA's standard, the PEL of asbestos exposure is 0.2 fibers/cm³ (fibers per cc). The Bioenvironmental Engineering team consists of five workers, four males (Worker #1, #3 #5, and #6) and one female (Worker #2), one supervisor, and a male bioenvironmental engineer (Worker #4). The team was charged to remove asbestos from Bldg. 6371 which was a two-story building with an attic and a basement, and previously used as a clinic. Among five workers, Worker #1 and #6 was in charge of actively removing the asbestos. Worker #2 and #3 were doing the general clean-up, while Worker #5 was doing either removal or clean-up depending on demand. Throughout the entire project all workers wore both the whole-body protective cloth and the respiratoratory mask.

The removal project was begun on March 28, 1991 and ended on September 5, 1991. The completion of the project was divided into three phases. Phase 1A was to remove the asbestos from the attic and part of the second floor which can not be kept from contamination of the removal work proceeded first in the attic. Phase 1B was to remove the asbestos from the remaining part of the second floor. Phase 2 was then to clean the asbestos from the first floor, while Phase 3 was destined to clean the basement.

Before actual removal of asbestos was started, personal samples were collected from Worker #1, #3, and #5 on two different days, March 28 and April 1. These exposure measurements are viewed as baseline control. Area baseline-control samples were collected on April 17. The actual removal of asbestos, namely Phase 1A, was carried out on May 9. Phase 1A was completed on May 30. Before

Phase 1B was started, Area samples were collected on June 5 from the north, south, east, and west side of the attic, old operating room, staircase, center hallway, and Room #1-#7 on the second floor. Phase 1B began on June 18 and ended on June 21. Again, before the start of Phase 2, area samples were collected on July 1 from Room #1-#11 and center hallway of the second floor, upstairs' staircase and downstair's dirty room of the first floor. Phase 2 was started on July 16 and ended on July 25. Once again, area samples were collected on August 1 from Room #1-#15, hallway #1-#5, the back staircase of the first floor, and the trailer's dirty room immediately attached to the first floor of the building. Phase 3 was then started on August 14 and ended on August 23. Another area samples were taken on August 29 from the storage area, X-ray room, old laboratory, hallway between B- and C-Wing, small room in C-Wing, Trailer's Hallway, and room inside containment, while on September 5 from X-ray room, little room and Hallway in C-Wing. The collection of area samples is to make sure that asbestos was indeed cleared from the building.

The data, $\{x_{c,i}\}_{i=1}^{n}$, obtained from personal samples of five workers and an bioenvironmental engineer are given in Table 1, in which the word 'Wet' means the filter cassette of personal monitor was wet and torn apart; hence it can not be measured and was treated as missing data not used in the analysis. To facilitate calculation, all dates in Table 1 are coded by integers from 1 to 29. Before the removal work was actually started, data served as baseline control were collected on Day #1 and #2. Neither these data nor the last observation ""none-detected"" of Worker #5 were used in the analysis. Also, since there is no way to measure the correct exposure level of asbestos from overloaded data, the asbestos concentration could be any number from 0 to 100 fibers per cc. Based upon personal engagement in the removal work, the bioenvironmental engineer (Worker #4) gave the author a conservative estimate of the asbestos concentration of overloaded data as 5 fibers per cc.

Result and Discussion

Since random sampling was not possible and the data were collected on consecutive workdays, the day-to-day sampling errors must be somewhat autocorrelated. Additionally, the removal work has three different phases, it seems possible that the worker's mean exposure level over the entire project period might have changed at a certain point in time, namely, the mean expouse level is not a single constant for the entire period. By plotting daily exposure levels of the (natural) logarithm-transformed data, $\{x_{\ell,i}\}_{i=1}^{n}$, of Worker #1 and #6, it fluctuates widely from day to day, as shown by the first plotted line of Figure 1. By perusing daily exposure levels, it is suspected that the mean exposure level might have changed, respectively, on Day #16 and #5 for Worker #1 and #6 Indeed, after applying a more powerful method of the rank cusum chart to the data,[5] it is confirmed that the mean exposure level had indeed changed for both workers at the above suspected days by noting that a significant change of the slope between successive data points did occur at the suspected day (2nd line of Figure 1). The fluctuation of daily exposure levels varies for other workers too. After applying the rank cusum chart, changes in their mean exposure levels for other workers were noticed. But, changes were not statistically significant at the level of $\alpha = 0.05$.[5] To save space, their graphs are not shown here.

Before taking changes in mean exposure level and the autocorrelated sampling errors into consideration, descriptive statistics under the assumption of lognormality and normality are given, respectively, in Table 1(a)-(b) when the overloaded data were used in the calculation. By examining the ratio of the largest and smallest observations and the coefficient of variation (CV), the assumption of lognormality appears more appropriate than the normality. Before using (7) to remove the autocorrelated sampling errors, the autocorrelation coefficient ρ has to be estimated from the data first. For Worker #1 and #6, the data were divided into two subsets, called the first and second phase, by their change-point dates as can be read from Figure 1 at which the slope between succesive points has changed from negative to positive, and then estimate ρ separately from the data of two subsets. To facilitate computation, the ARIMA procedure of the SYSTAT[12] was employed and the point-estimates $\hat{\rho}$ for all workers are given in the fifth column of Table 3. Next, substituting $\hat{\rho}$ into (7) and then calculates s_{ℓ} of

(8). Incidentally, by assuming that the error variances are the same for both the first and second phases, the pooled estimates of s_ℓ for Worker #1 and #6 are 2.33 and 2.41, respectively. Note that the geometric standard deviation, $\hat{\mu}_g = \exp(s_\ell)$, changes dramatically for all workers (comparing the 4th row of Table 2(a) and the 4th column of Table 3). This shows that the autocorrelation has a great influence on computing exposure variability. Hence, when dealing with exposure time series data, if the autocorrelated sampling errors are not removed, a straight calculation of geometric standard deviation from the (natural) logarithm-transformed data, $\left\{ x_{\ell,i} \right\}_{i=1}^{n}$, is erroneous.

Before (6) can be employed to test compliance for each worker, both PEL_ℓ and $s_{\bar{x}_\ell}$ have to be computed. Using (4) with PEL = 0.2 fibers per cc and s_ℓ, the computed values of PEL_ℓ are given in the second row of Table 4. By substituting the estimated values of $\hat{\rho}$ and s_ℓ into (5), the obtained standard errors, $s_{\bar{x}_\ell}$, are given in the sixth column of Table 3. Finally, using $\hat{\mu}_\ell = \bar{x}_\ell$, (2nd row of Table 2(a)), PEL_ℓ, (2nd row of Table 4), and $s_{\bar{x}_\ell}$, (6th column of Table 3), the computed t-values of (6) are given in the third row of Table 4. By comparing with the critical values of t-distribution with n - 2 degrees of freedom, (4th row of Table 4), Worker #1, #3 and #5 are shown to be in-compliance with the OSHA's standard at the 5% significance level, since the computed t-value is less than the corresponding critical value of t-distribution.

Ratios of the average exposure level estimated from the data to the OSHA's PEL for all workers are given in the sixth row of Table 4. It is worthy of noting that for Worker #1 and #3 their average exposure levels are 2.5 and 3.7 times the OSHA's PEL. However, after removing the autocorrelated sampling errors and taking the longitudinal exposure variability into consideration, our method shows that their average exposure levels are still in compliance with the OSHA's standard.

If the overloaded data were not used in the analysis, analysis remains unchanged for Worker #5 since he has no overloaded data, while Worker #4 merely had two observations left and insufficient for statistical analysis; hence deleted from further analysis. For other workers, we recalculate everything as given in Tables 5, 6 and 7. Now, clearly, both the estimated mean exposure level and the associated variability became smaller. Yet all of our analysis and remarks about the importance of removing autocorrelated errors and detecting changes in the mean exposure level remain valid (Figure 2). For the later data set only Worker #2 has her average exposure level 1.8 times the OSHA's PEL. Nevertheless, by our method it was shown that her average exposure level (Worker #2) is still in compliance with the OSHA's standard.

To examine the overall longitudinal and cross-sectional exposure variability, the GLM procedure of SAS[11] was used to facilitate calculation from the transformed data of (7). The results with or without Worker #4 to be included in the ANOVA model are given, respectively, in the second and third rows of Table 8. If Worker #4 is included in the model, the longitudinal and cross-sectional exposure variability, (52% and 48%), are comparable (2nd row of Table 8). However, the longitudinal variability, (68%), is two times the cross-sectional variability, (32%), if Worker #4 is not included in the model. Since Worker #4 only has three observations, the inclusion of him in the model artificially inflates the cross-sectional variability. Therefore, the model without Worker #4 seems more reasonable. Furthermore, if both the overloaded data and Worker #4 are not used in the analysis, the longitudinal variability, (89%), is increased to 8 times the cross-sectional variability, (11%), as shown by the fourth row of Table 8. However, due to dropping the overloaded data, it might artificially inflate the longitudinal variability because the sample size for each workday becomes too small (Table 1). Also, only 43% and 61% of the total exposure variability are accounted for both longitudinal and cross-sectional exposure variability depending on whether the overloaded data were used in the analysis or not (5th column of Table 8). This implies that about 40% of the total exposure variability were still not accounted for some unknown factors. In addition, through the use of the same ANOVA model it was shown that none of the differences in the mean exposure level for the factors of Day-to-Day, Worker, Sex, and Weekday is statistically significant at the the level of $\alpha = 0.05$.

Conclusions

In recognition of the exposure variability a new statistical method is proposed for testing whether it will still be in-compliance with the OSHA's standard if the Worker's average exposure level is several times the OSHA's PEL. The asbestos exposure data is used as an illustration. This data set, consisting of personal samples from a team of five workers exposed to asbestos, possesses several interesting features. Fisrt, the mean exposure level for some workers has changed during the exposure period. A rank cusum chart was applied to detect changes in the mean exposure level. Second, the sampling errors for all workers are highly autocorrelated. An autoregressive transformation was applied to remove the autocorrelated errors from the data. It was noticed that if the autocorrelated errors were not removed, a straight calculation for the exposure variability could be erroneous. Third, it was shown that even if the Worker's average exposure level is several times the OSHA's PEL, it could still be in-compliance with the OSHA's standard. Finally, through the use of the ANOVA model, it was shown that the longitudinal exposure variability is the dominating one which could range from 2 to 8 times the cross-sectional variability. However, the longitudinal exposure variability might be artificially inflated because too few samples were available for each workday. To obtain a more definite relationship between the longitudinal and the cross-sectional variability, more samples need to be collected from more workers.

Acknowledgments

REFERENCES

1. Box, G.E.P. and G.M. Jenkins: Time Series Analysis: Forecasting and Control. 2nd edition, Holden-Day, San Francisco, CA (1976).

2. Cochran, D. and G.H. Orcutt: Application of least squares regression to relationship containing autocorrelated error terms. J. Am. Stat. Assoc. 44: 32-61 (1949).

3. Esman, N. and N. Hammad: Lognormality of environmental sampling data. J. Environ. Sci. Hlth A12: 29-41 (1977).

4. Francis, M.,S. Selvin, R.C. Spear, and S.M. Rappaport: The effect of autocorrelation on the estimation of workers' daily exposures. Am. Ind. Hyg. Assoc. J. 50: 37-43 (1989).

5. Lee, T-S.: Rank cusum test for detecting changes in mean. The 26th Symposium on Interface between Computing Science and Statistics: (to appear) (1994).

6. Leidel, N., K. Busch, and J. Lynch: Occupational exposure sampling strategy manual. USDHEW, NIOSH Publ. No. 77-173. National Institute for Occupational Safety and Health, Cincinnati, Ohio (1977).

7. Nicas, M, B.P. Simmons, and R.C. Spear: Environmental versus analytical variability in exposure measurements. Am. Ind. Hyg. Assoc. J. 52: 553-557 (1991).

8. OSHA: "Industrial Hygiene Field Operation Manual". OSHA Instruction CPL 2-2.20, April 30, (1979).

9. Rappaport, S.M.: The rules of the game: An analysis of OSHA's enforcement strategy. Am. J. Ind. Med. 6: 291-303 (1984).

10. Rappaport, S.M.: Assessment of long-term exposures to toxic substances in air. Ann. Occup. Hyg. 35: 61-121 (1991).

11. SAS: Statistical Analysis System-SAS User's Guide: Statistics Version 5 Edition. SAS Institute Inc., Box 8000, Cary, NC 27511-8000 (1985).

12. SYSTAT: Statistical Applications. SYSTAT, Inc., Evanston, IL (1987).

Figure 1. The plot of daily exposure levels and the rank cusum charts for Worker #1 and #6 when the overloaded data are used in the analysis.

Asbestos Exposure of Worker 1

Asbestos Exposure of Worker 6
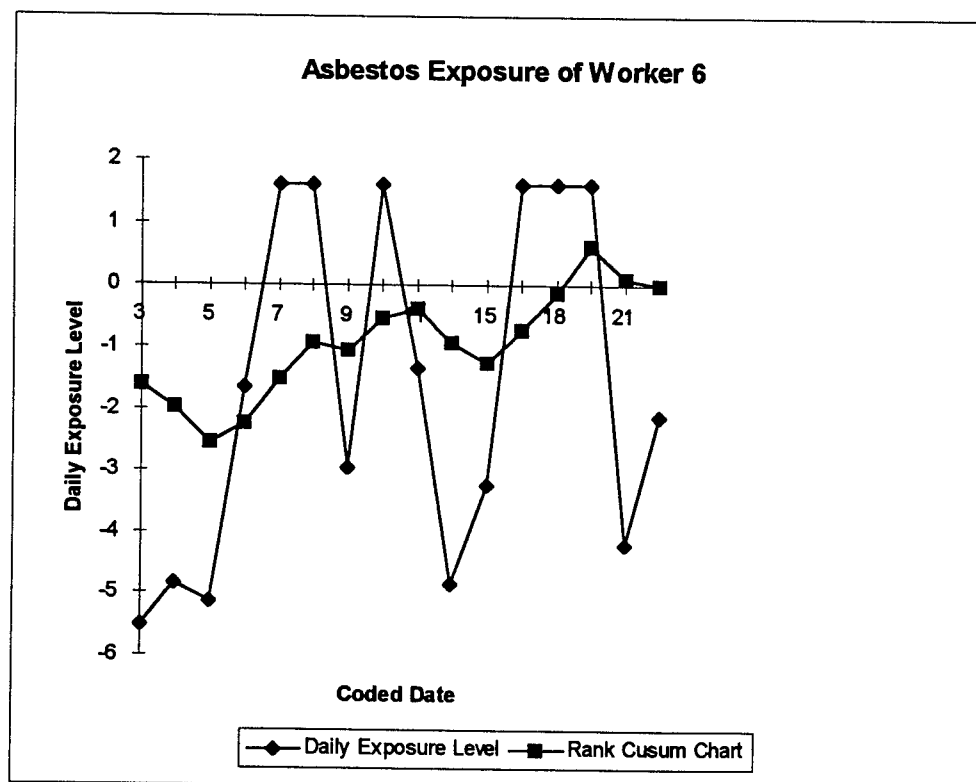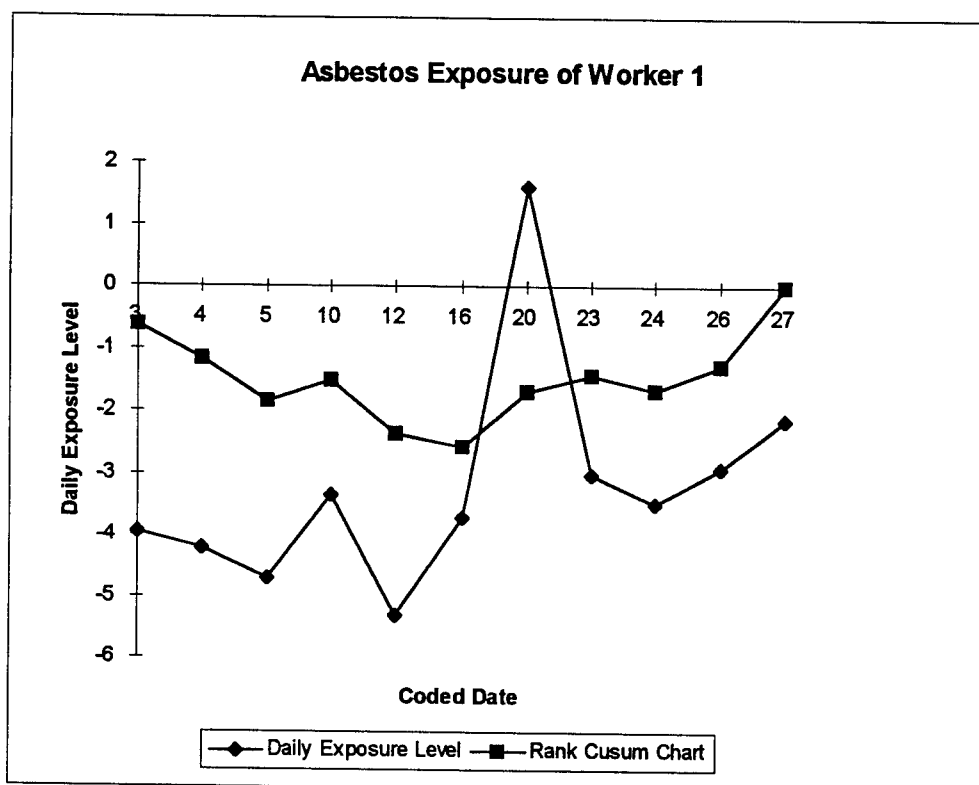
Figure 2. The plot of daily exposure levels and rank cusum charts of Worker #1 and #6 when the overloaded data are not used in the analysis.

Asbestos Exposure of Worker #1
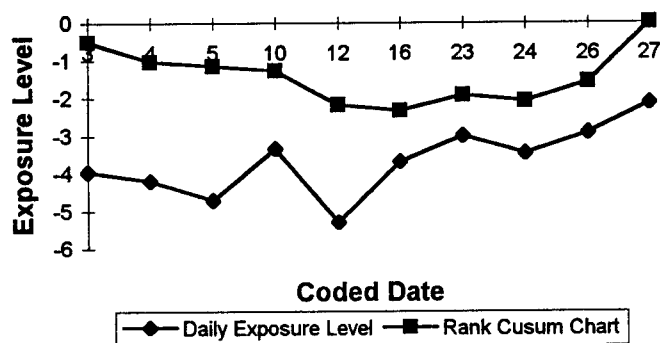
Asbestos Exposure of Worker #6
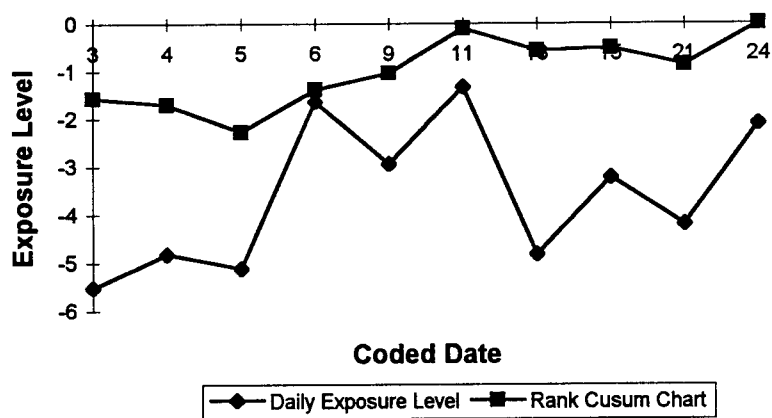
Table 1 The day-to-day asbestos exposure of six workers at Lackland AFB.

Date (Weekday)	Worker					
	#1	#2	#3	#4	#5	#6
3-28-91 (Th)	.003				.007	
4-1-91 (M)	N/D[a]		N/D		N/D	
5-9-91 (W)	.019		.027	.012		.004
5-10-91 (Th)	.015					.008
5-13-91 (M)	.009		.006			.006
5-14-91 (T)	Wet[b]	1.102	.02	.553		.194
5-15-91 (W)		1.915	Wet			5.0*
5-21-91 (T)		.071		5.0*		5.0*
5-24-91 (F)					.03	.053
5-28-91 (T)	.035				.027	5.0*
5-30-91 (Th)		.069				.264
6-18-91 (T)	.005		.003		.003	
6-20-91 (Th)			.006		.004	.008
6-21-91 (F)			.014		Wet	
7-16-91 (T)					.043	.04
7-17-91 (W)	.025					5.0*
7-18-91 (Th)			5.0*			
7-19-91 (F)		.008	Wet			5.0*
7-22-91 (M)			5.0*		.006	5.0*
7-23-91 (T)	5.0*	5.0*	.001			
7-24-91 (W)						.015
7-25-91 (Th)			.009		.018	
8-14-91 (W)	.05					
8-15-91 (Th)	.031	.081	.094			.121
8-16-91 (F)		.142			.004	
8-19-91 (M)	.054				.059	
8-21-91 (W)	.118	.069	.133		.114	
8-22-91 (Th)		.002	.1			
8-23-91 (F)		.111	.002		N/D	

[a] N/D denotes none detected.

[b] Wet denotes that the filter was wet and torn apart; hence it can not be measured.

*This is an assumed value because filter cassettes were overloaded.

Table 2 Summary statistics for workers exposed to asbestos when the overloaded data were used in the analysis.

(a) The assumption of lognormal distribution

Statistics	Worker					
	#1	#2	#3	#4	#5	#6
$\hat{\mu}_g$	0.04	0.12	0.03	0.32	0.016	0.20
$\hat{\mu}_\ell$	- 3.19	- 2.12	- 3.51	- 1.14	- 4.13	- 1.63
s_g	6.1	9.7	13.7	21.2	3.6	17.3
s_ℓ	1.82	2.27	2.62	3.05	1.28	2.85
Minimum	- 5.30	- 6.21	- 6.91	- 4.42	- 5.81	- 5.52
Maximum	1.61	1.61	1.61	1.61	- 2.17	1.61
CV	57%	107%	74%	268%	31%	174%
sample size	11	11	14	3	10	16

(b) The assumption of normal distribution

Statistics	Worker					
	#1	#2	#3	#4	#5	#6
Mean ($\hat{\mu}_c$)	0.49	0.78	0.74	1.86	0.03	1.92
Standard deviation	1.50	1.53	1.80	2.74	0.034	2.47
Minimum	0.005	0.002	0.001	0.012	0.003	0.004
Maximum	5.0	5.0	5.0	5.0	0.114	5.0
CV	307%	196%	242%	148%	113%	128%
sample size	11	11	14	3	10	16

Table 3 The standard error of the sample mean adjusted for the effect of autocorrelation when the overloaded data were used in the analysis.

Worker	Phase	n	$\hat{\sigma}_g (= \exp(s_\ell))$	$\hat{\rho}$	$s_{\bar{x}_\ell}$	
					Separate	pooled
#1	First	6	4.1 (=exp(1.42))	0.94	1.63	1.92
	Second	6	19.4 (= exp(2.97))	0.68	2.18	
#2	Single	11	13.6 (=exp(2.61))	0.49		1.26
#3	Single	14	21.1 (=exp(3.05))	0.74		1.85
#4	Single	3	1.0 (= exp(0.023))	0.74		0.02
#5	Single	10	5.5 (= exp(1.71))	0.91		1.46
#6	First	3	2.0 (= exp(0.67))	0.96	0.85	1.20
	Second	14	12.2 (= exp(2.50))	0.16	1.25	

Table 4 Testing for compliance of six workers exposed to asbestos when the overloaded data were used in the analysis.

Empirical Statistics	Worker					
	#1	#2	#3	#4	#5	#6
PEL_ℓ	- 4.32	- 5.02	- 6.26	- 1.61	- 3.07	- 4.51
Computed t-value	0.58	2.30	1.49	23.7	- 0.72	2.40
Critical value of t-test at $\alpha = 0.05$	1.83	1.83	1.78	6.31	1.86	1.76
In-Compliance	Yes	No	Yes	no	Yes	No
$\hat{\mu}_c / PEL$	2.5	3.9	3.7	9.3	0.2	9.6
Excursion rate	1/11	3/11	2/14	2/3	0/10	7/16

Table 5 Summary statistics for workers exposed to asbestos when the overloaded data were not used in the analysis.

(a) The assumption of lognormal distribution

Statistics	Worker				
	#1	#2	#3	#5	#6
$\hat{\mu}_g$	0.025	0.082	0.013	0.016	0.028
$\hat{\mu}_\ell$	- 3.67	- 2.5	- 4.36	- 4.13	- 3.57
s_g	2.51	7.5	4.9	3.6	4.6
s_ℓ	0.92	2.01	1.59	1.28	1.53
Minimum	- 5.30	- 6.21	- 6.91	- 5.81	- 5.52
Maximum	- 2.14	0.65	- 2.02	- 2.17	- 1.33
CV	25 %	80 %	36 %	31 %	43 %
sample size	10	10	12	10	10

(b) The assumption of normal distribution

Statistics	Worker					
	#1	#2	#3	#4	#5	#6
Mean ($\hat{\mu}_c$)	0.036	0.357	0.035	1.86	0.03	0.071
Standard deviation	0.03	0.64	0.046	2.74	0.034	0.092
Minimum	0.005	0.002	0.001	0.012	0.003	0.004
Maximum	.118	1.915	0.133	5.0	0.114	0.264
CV	91 %	179 %	134 %	148 %	113 %	128 %
sample size	10	10	12	3	10	10

Table 6 The standard error of the sample mean adjusted for the effect of autocorrelation when the overloaded data were not used in the analysis.

Worker	Phase	n	$\hat{\sigma}_g (= \exp(s_\ell))$	$\hat{\rho}$	$s_{\bar{x}_\ell}$	
					Separate	pooled
#1	First		4.1 (=exp(1.42))	0.94	1.63	1.26
	Second	5	1.8 (= exp(0.56))	0.88	0.51	
#2	Single	10	8.9 (=exp(2.19))	0.76		1.53
#3	Single	12	7.2 (=exp(1.98))	0.96		1.83
#5	Single	10	5.5 (= exp(1.71))	0.91		1.46
#6	First	3	2.0 (= exp(0.67))	0.96	0.85	1.27
	Second	8	7.5 (= exp(2.02))	0.68	1.37	

Table 7 Testing for compliance of six workers exposed to asbestos when the overloaded data were not used in the analysis.

Empirical Statistics	Worker				
	#1	#2	#3	#5	#6
PEL_ℓ	- 2.14	- 4.01	- 3.57	- 3.07	- 2.52
Computed t-value	- 1.21	0.99	- 0.43	- 0.72	- 0.83
Critical value of t-test at $\alpha=0.05$	1.86	1.86	1.81	1.86	1.86
In-Compliance	Yes	Yes	Yes	Yes	Yes
$\hat{\mu}_c / PEL$	0.2	1.8	0.2	0.2	0.4
Excursion rate	0/10	2/10	0/12	0/10	1/10

Table 8 A comparison on the longitudinal and cross-sectional variability of the asbestos exposure.

Model	Exposure variability			R^2
	Longitudinal (%)	Cross-sectional (%)	Total	
Worker #4 was included	5.3 (52%)	4.8 (48%)	10.1	0.46
Worker #4 was excluded	5.2 (68%)	2.4 (32%)	7.6	0.43
Both overloaded data and Worker #4 were excluded	3.4 (89%)	0.4 (11%)	3.8	0.61

A STUDY OF INTERACTION IN DISTANCE LEARNING

Robert G. Main, Ph.D.
Professor
Eric Riise
Graduate Student
Department of Communication Design

California State University, Chico
Chico, CA 95928

A STUDY OF INTERACTION IN DISTANCE LEARNING

Robert G. Main, Ph.D.
Professor
Eric Riise
Graduate Student
Department of Communication Design
California State University, Chico

Abstract

Interaction in distance learning was studied. A survey of the literature found that most studies were lacking in rigor and the methodologies were weak in regards to interaction. To answer the many questions about interaction effects in distance learning, a better definition of the variable interaction is needed. This paper lays out a taxonomy of interaction for evaluation and research.

A STUDY OF INTERACTION IN DISTANCE LEARNING

Robert G. Main, Ph.D.
Eric Riise

Introduction

In the past distance learning has been largely used to bring education and training programs to learners who would otherwise not have access to the classes offered (*U.S. Office of Technology Assessment, 1989*). Courses and programs have been offered primarily for adult learners interested in college credit or vocational and professional improvement. Distance learning networks are operated for the most part by colleges and universities or by corporations and government agencies.

There is ample evidence in the literature that, in most cases, distance learning appears to be as effective as face-to-face instruction in the classroom (Moore, 1989). A comprehensive review of the current research regarding the use of dynamic video media in instruction conducted by Wetzel, et al, (1994) found, 'The general conclusion from this evolving field is that it is possible to have no decrement or only a small decrement at a remote site compared to the performance of students at the live transmitting site, (p. 20)". They examined studies of both preproduced telecourses (i.e., non-interactive) and interactive teletraining in terms of effectiveness, acceptance and costs. They found the primary attraction of distance learning for students is convenience: proximity to where they work and live and flexibility in personal scheduling and work requirements.

In a comprehensive review of the current research regarding the use of dynamic video, the Texas Higher Education Coordinating Board (1986) reviewed the course results of four college level telecourses and found student achievement was comparable to conventional on-campus classes. They also examined grade distributions of Texas community colleges offering telecourses and found they did not differ significantly from traditional classroom grade spreads. A review of the literature by Miller, et al, (1993) could not identify a single study that has shown distance learning diminished content learning. Some studies found advantages of distance classes over traditional classroom instruction (e.g., Barron, 1987, Weingard, 1984, and Keane and Cary, 1990).

Most distant learners report they are satisfied with their remote instruction and some reportedly preferred the distance learning mode. However, the limited number of studies in the

affective domain and the lack of rigor in the methodologies do not permit reliable conclusions about preference.

. That these effects hold true for many subjects and a variety of media and delivery means, indicates that learner motivation may be an overarching factor in the learning process *for these students*. Indeed, it has been asserted that motivation is the single most important factor for student learning and when motivation is high, it is difficult to prevent learning (Main, 1992).

However, distance learning is entering a new phase. The transformation from analog to digital communication technology is creating a new environment for multimedia interactive instructional media and telecommunication networks. Distance learning is no longer being viewed as simply a means to provide access for those unable to meet in the classroom, but as a viable alternative to classroom instruction as a primary mode of instruction . The promise of two-way interactive video, voice and data available in every home and office via the information superhighway has fired the imagination of educators and non-educators alike in the potential for providing elementary and secondary schooling in classrooms without walls. Decisions are now being addressed on the basis of how cost-effective distance education is compared with traditional classroom instruction. The sentiment is reflected in a recent comment by a corporate officer that the "lean budgets of today's economy drive alternative training and educational delivery systems. Traditional stand-up instruction does not stand up to the scrutiny of the cost conscious business manager." (Grant, 1994) Universities are experimenting with delivery of instruction to students in their dormitory rooms or homes through local area networks or public data services such as America Online. Public telephone and cable distribution systems are under study as well, as a means of providing instruction without assembling students in a classroom.

Whether the success of distance learning with adult learners will work equally well for all students is still a question. Most applications to date have been involved with academically advanced high school students and independent adult learners--individuals who presumably already possess strong study skills, high motivation and discipline (*U.S. Office of Technology Assessment, 1989*). The Congressional Office of Technical Assessment (OTA) has concluded that research of distance learning for elementary and secondary application would be most usefully concentrated on practical questions about the educational experience such as learner outcomes of various teaching techniques and instructional design approaches.

With distance learning being considered as a replacement for traditional classroom, the designers and developers of distance learning instruction can no longer depend upon the intrinsic

motivation of self-selection. Like the public school teacher in the traditional classroom, the students will present an array of interest in the subject and education for the distance learning instructor. The changing nature of the distant learner from adult volunteer to adolescent required attendee presents new requirements for the instructional designer and teacher.

This study examines interactivity as a function of the instructional design and presentation of distance learning lessons. The complex interplay of interaction in distance learning is not well understood at this time (Haynes and Diehorn, 1992). One reason is the relative dearth of studies examining interaction in distance learning education. Others include the poor controls used in the research that has been conducted and a reliance on self-selected groups exposed to the distant learning and traditional classroom conditions. Finally, there is the relatively simplistic manner in which interaction has been defined in the studies. It is also likely that studies showing poor performance for distance learning situations are less likely to be submitted for publication or published when they are submitted.

Background

Intuitively we know that interaction is important in the instructional process. We strive for interaction in the traditional classroom. The concept of small teacher-student ratios is based on the belief that the smaller class size permits a richer interaction. The ultimate learning environment is considered to be one-on-one where the instruction can be individualized to the student's perceived needs and learning style. It is axiomatic that proximity in interpersonal communication enriches interaction. Wetzel, et al, (1994) determined that, "Increasing the degree of fidelity or interactivity of video teletraining to that with live instruction generally increases effectiveness and satisfaction "(p. 21). But the empirical evidence is weak and the studies cited are generally lacking in methodological rigor. Klinger and Connet (1992) state, "...telecourses must include a strong element of interaction to be truly effective as a learning method. Interaction is *essential* for the student to remain interested and steered forward for success" (p. 88). Their conclusions, however, are based on experience rather than empirical studies. How, then, are we to explain the results of the many studies which indicate there is little difference in learning between students in the traditional classroom and students at distant learning sites?

It is difficult to tell from the literature. There are very few studies that have examined interactivity as an independent variable and those that purport to have studied its effects generally looked at interaction only in terms of frequency. For example, Van Haalen and Miller (1994) reported on interactivity as a predictor of student success in a satellite learning program, but interactivity was measured on the basis of telephone logs recording only the number of calls from

students to the teacher both during and after class for the school year. No attempt was made to capture the length of the interaction or its topical relevance. Most interaction reports are observational and associated more with learner attitudes about the delivery mode than with achievement. Rupinski (1991), for example, found that student preferences for traditional classroom training can be reduced by making conditions at the remote site (including two-way video) more like a "live" classroom.

From studies where interaction is included as a variable, the effects of interaction of learning outcomes is ambiguous. In a study comparing instruction by audiotape, videotape and telelecture, Beare (1989) found the lack of individual opportunity to interact with the instructor regularly did not significantly reduce student scores on course examinations. In a one-way video course with two-way audio, Van Haalen and Miller (1994) reported interaction effects were not linear but, rather, a polynomial curve in the form of an inverted U. At each end of the interactive continuum, student performance (in terms of course grades) was poor. They only measured student initiated interactions with the instructor, however, and not interactions designed into the instructor's presentation as student-student discussion activities. It is possible that in this situation, the students with the most questions are those with a need for additional information to keep up with the instruction. Conversely, students who never ask questions may be reluctant to expose a lack of knowledge.

A problem with the descriptive studies is the lack of a control group. How do the students in the distant learning class differ from those in the residential classroom? Zigerell (1986) gives a hint with his survey of telecourse students in community college courses. Most of the students had not taken a telecourse before (65 percent). Of those, 69 percent were women, and carried less than 10 semester units. About half worked at least 40 hours a week. Only 17 percent said they were enrolled because they preferred distance learning. This is quite different from most residential college students.

In one of the more carefully designed comparative studies, (Simpson, et al, 1991) found the most critical condition for success in interactive teletraining is the ability of students to see the instructor and have two-way audio communication . Two-way video appeared to have little effect, but any degradation in audio quality caused negative comments. Not surprising since most instruction is still language-based. In comparing final examination scores, the decrease between student performance at the originating site and the remote sites was less than three percent in any of the instructional modes. The value of Simpson's studies are that they compared complete courses-

-not just modules covering a few hours of instruction. Stoloff (1991) found distant learners became more indifferent to differences between teletraining and traditional classroom methods of instruction over time, but instructors still favored their classroom.

There is a need in distance learning research to adopt an expanded view of effective teacher-student communication. It involves integrating a variety of communication forms and channels that include verbal communication, vocal communication--the volume, rate, tone, pitch and inflection--mediated messages, body language and situational messages--manipulation of distance, time and number of participants (Hennings, 1975).

Instruction, Communication and Interaction

Teaching is primarily a communication art. If we accept the interdependent relationship between source and receiver in the communication process described by Berlo (1960), then teaching should emphasize interaction between instructor and learners. We learn by taking an active role in the process (Hefzallah, 1990). Buckminster Fuller asserts the instructional environment "is an interacting situation in which the continuity of experience and the relating of experience are critically important." (1966, p 16). Hefzallah summarized the connection succinctly, "to teach is to communicate, to communicate is to interact, to interact is to learn (1990, p 38)."

Socrates knew the value of interaction in learning. Students were required to discover knowledge through a series of questions and answers. By contrast, the Sophists were the first lecturers. They knew everything and were ready to explain it (Highet, 1957). But, here's the rub. While the Sophists grew rich, dressed in royal purple and traveled by sedan chair, Socrates' sandals were worn and his tunic undyed. His discovery learning was not cost-effective. To make intelligent decisions in designing distance learning systems and lessons, we need information about the trade-offs between effectiveness and efficiency in the amount and quality of interaction in the instructional process.

The seminal studies by Chu and Schramm (1979) established that children learn efficiently from instructional television and from instructional radio "given favorable conditions" (p.vi and 1). The favorable conditions generally refer to the similarities in presentation where students in both groups hear the same lecture, see the same visual and read the same printed materials. Most of the studies supportive of these conclusions contrasted students in traditional classes with those at remote locations. For the most part, the studies reviewed by Chu and Schramm used a mass media model for the instruction, i.e., the transmission was largely one-way with feedback limited and delayed. In this mode, the student is largely passive, at least in terms of real-time interpersonal

interaction. This is the conduit theory of communication applied to distance learning described by Clark (1983) as an analogy in which "media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition," (p. 445). To extend Clark's analogy, however, if the truck can't be refrigerated to carry fresh fruit, that does affect our nutrition. So, the capability of the delivery system is a factor in the instruction presentation and possibly student learning.

In his review of the distance learning literature, Kozma (1991) acknowledged the mass media conduit theory, but concluded the interpersonal theory of communication with its rich and immediate feedback was more appropriate for the interactive nature of teaching. Distance learning should attempt to replicate the "live" classroom through "virtual medium" (Kozma, 1992, p. 182).

It's not only the quantity of interaction that affects the learning . The quality of interaction is also a factor. The interminable prompt to press ENTER that was so common in early computer aided instruction although interactive was a numbing experience. Dale states "education has to choose creative interaction of the learners over rote imitative reaction (1978, p 24)." This is what designers of distance learning education face: determining the amount and nature of interaction that is most effective and efficient for achieving the learning objectives of the class.

Two issues are identified by Miller, et al (1993) as being important in considering how well distance learning duplicates the learning environment of the traditional classroom. The first is whether interaction is reduced among distance learners. Even though the technical capacity for such interaction is available, students may be inhibited from participating interactively by a variety of reasons such as awkwardness in interrupting or unfamiliarity with equipment. The second issue concerns the degree of student engagement in traditional and distance learning conditions, i.e., do distant students tend to be less attentive in a distance learning environment? Nadel (1988) in a comparative study of distance learning modes concluded that students learn from any medium, in school or out, whether they intend to or not, providing the content of the medium leads them to *pay attention* (emphasis added). That is a very large proviso and corresponds to Kozma's (1991) concept of involvement which is manifested behaviorally as participation and can be measured by interactions.

A Taxonomy of Interaction for Evaluation and Research

To answer the many questions about interaction effects in distance learning, a better definition of the variable interaction is needed. Interaction in distance learning is obviously a complex behavior. What is needed is a model for examining its many dimensions. Boak and Kirby

(1989) developed the System for Audio Teleconferencing Analysis (SATA) instrument for analyzing classroom interaction that affords some direction for researchers. Their schema has three categories: who initiates the interaction (student or instructor); the direction of the interaction (an individual student, the class as a whole, or instructor); and the context of the interaction (procedural, content specific, or social).

This schema can be expanded to examine more thoroughly the interaction process in the distant learning classroom. Six categories of interactivity have been identified by this researcher as relevant for distance learning research. These may not be comprehensive but provide a beginning point in developing a taxonomy of distance learning classroom interaction. They are **AMOUNT, TYPE, TIMELINESS, METHOD, SPONTANEITY**, and **QUALITY**. Each of these components are compound variables themselves with several levels.

1. There are two dimensions in measuring the **amount** of interaction--the frequency of occurrence and the length of the dialog. Frequency is perhaps the most commonly captured data in distance learning studies involving interaction. It is most often examined in terms of how often student feedback occurs, i.e., the mean occurrence per student per class period. Frequency can also be measured by how it is spaced during the presentation (clustering by instructional activity.) The length of each interaction is relatively straight forward and is of greatest interest when related to **type, method** and **quality.**

2. The **type** of interaction refers to the participants. In a distance learning class this would include instructor-student exchanges, student-student interactions, and student involvement with the lesson materials. The instructor-student interaction can be further categorized as to whether it is instructor initiated or student initiated. Student-lesson materials interaction may be either required or student choice. Each of these levels can be further subdivided as occurring within the class period or outside the class period. The measurement would be frequency of occurrence. The cells would appear as follows:

Fig. 1 Type of interaction

Initiated by		Student-	Student-lesson Materials		
Instructor	Student	Student	Required	Voluntary	
					Within Class
					Outside Class

Measuring student-to-student interaction will require some means for observing or recording activity at the distant learning site(s). Haynes and Dillon (1992) found distance students in a library science course interacted much less with the instructor and much more with each other during class even though they complained at times that it interfered with attending to the instructor. The results of the study did not indicate a significant difference in student performance between exam scores of distant and on-site learners which would seem to show interaction type has little effect on distance learning. There are some methodological problems with the study, however, that make generalization of the findings difficult.

3. **Timeliness** is a measurement of the immediacy of the communication feedback. It is the amount of time between the attempt to interact begins until the message is received by the addressee. It is an indicator of the efficiency of the communication system. It presumes a two-way communication system. In those situations where the instruction is preproduced and delivered on schedule or on demand, there is no interaction in the class. Broadcast television and cable delivered instruction fall into this category. It's the mass media model of communication. An example might be Ken Burns' *Civil War* television series aired over PBS. It was certainly educational, but a passive one-way delivery when viewed on PBS. However, when video tapes of that series are used by a teacher as instructional material in a distance learning history class where feedback is expected of students discussing the programs, the instruction becomes interactive. Timeliness is a continuous variable that ranges from zero in the real-time interactions of a traditional classroom to several days or even weeks for a correspondence course administered through the post office.

4. The **method** of interaction refers to the manner in which the communication message is encoded. Voice is the most common method of interaction in the traditional classroom. Satellite transmissions of one-way video with a two-way telephone audio channel have been the system of

choice for most distance education and training systems. However, there has been considerable interest in text-based interaction systems using computer-based data network delivery. With the conversion of analog to digital communication and the interest in establishing high capacity public switched digital infrastructure, there is an expanding effort to determine how this information superhighway can be exploited for education and training use. Already, compression technologies allow two-way, real-time digital video and audio transmission over conventional twisted pair phone lines through digital switches albeit without full motion or the fidelity of analog television. The **method** of interaction, then, should be addressed in studies of interaction effects. In addition to voice and text, interaction may occur through visual non-verbal gestures, response pads, graphic display, and photos. There are obviously many combinations and sub-levels possible with the various methods that need to be considered in developing the measurement methodology and instrument especially when newer multimedia workstations are used in the delivery of instruction.

5. The **spontaneity** dimension of interaction refers to whether the feedback is a planned event embedded in the lesson plan as part of a learning activity or a spur of the moment exchange triggered by the presentation. It may be important to determine whether *ad hoc* interactions are one-on-one or part of a group discussion. **Spontaneity** can be cross-tabbed with **amount, type** and other variables of the interaction schema.

6. The **quality** of interaction is the most difficult dimension of interaction to define operationally. The possible levels are almost infinite. Many of the other categories have quality implications and a case could be made that this is an overarching variable that subsumes all the components of interaction. For purposes of this taxonomy, **quality** is defined in five dimensions: *intensity, relevance, depth, formality,* and *opportunity*. *Intensity* reflects the emotional involvement of the participants in the interaction. The levels are routine (which includes repetitive, procedural and expected responses); interested (exploratory, explanatory, and expansive), and emotionally involved (excitement, fear, enjoyment, attachment, and anger). It is difficult to distinguish the intensity of a communication exchange, but trained observers can discriminate among the categories.

The components of *relevance* are classified as professionally related, involve the lesson content (subject matter) or have personal relevance. *Depth* is a continuum ranging from the trivial to substantive. The *formality* of the interaction is classified as formal or informal. *Opportunity* is the ability to interact when desired. It could be a function of class size, the technical capability of the system, or the instructional design of the lesson that accommodates interactions. Real-time two-way audio and video is expensive and the cost increases in direct relation to the number of distant

learning sites. It is important in emulating a traditional classroom, but the value decreases as the number of students in the class increases. The *opportunity* for interaction is inversely correlated with the class enrollment. The effect of class size is as true for the student in the traditional classroom as it is for the distant learner. Everyone remembers those large lecture halls where a professor addresses a class of hundreds. A satellite broadcast may enlarge the class to thousands of students. The chance of interacting with the instructor dwindles no matter how sophisticated the communication system. While the concept of the President appearing on a national talk show to interact with the public is politically appealing, the opportunity of any particular citizen to actually ask the president a question (never mind a give and take dialog) approaches the probability of winning the lottery. The idea that the caller who does get to ask a question represents some number of other viewers or listeners may have some validity, but is accomplished more economically by use of studio questioners. It is worthwhile measuring timeliness, however, even when the size of the class makes opportunity difficult. A study by Fulford and Zhang (1993) suggests the perception of overall interaction is a stronger predictor of student satisfaction than personal interaction. Although the class size was only 123 students in five locations, two of which had one-way video and two-way audio and three with two-way video and audio. The perception of overall interaction (self report) and satisfaction with the class had a strong correlation despite the actual number of personal interactions. The strength of "vicarious" interaction effect did diminish from the first of the three sessions to the last. We shouldn't be too surprised by these findings. The appeal of game shows and talk shows is largely the interaction between host and guests or contestant.

This taxonomy of interaction variables provides a framework for research and evaluation of the effects of interaction in distance learning. It may need modification and elaboration as new questions arise, but it allows the research to proceed more systematically in order that findings may be grouped for meta-analysis and meaningful comparisons made among studies. The next step is to develop operational definitions and measurement instruments for each variable that can be tested for accuracy and validity. The goal is to establish a body of literature from which theoretical concepts and generalizations can be made as to the efficacy of interaction activities that will be useful to system designers and instructional developers of distance learning instruction. The need for better methodology in distance learning studies is apparent. Research to date indicates there is little difference in achievement attributable to delivery technique. Intuitively that does not seem right even though studies have consistently reported performance of standardized tests to be similar, regardless of medium used (Salomon and Clark, 1977 and Ritchie and Newby, 1989).

<u>Relating Interaction with Other Variables</u>

Interaction always occurs within a context. The utility of organizing the dimensions of interaction variables lies in finding how they relate to other components of distance learning. There are numerous factors that may be affected by, or have an affect on, interaction in distance learning. Generally these factors can be classified as those concerned with the course and those dealing with its delivery, i.e., the communication technology. The communication technology is continually changing and especially at this watershed stage of conversion from analog to digital communication. Not only are the media becoming amorphous with digital multimedia, but also the industry infrastructure is in a state of flux as telephone, cable, and television companies seek acquisitions, alliances and mergers that will position them as players in the digital, interactive, multimedia future of the information superhighway. New products, new systems and new capabilities will demand continuing research for its effects on distance learning interaction.

Instructional strategies and activities involve all the components of instructional design with the added complexity of distance delivery (Wagner, 1990). There is a large body of literature available on instructional process, but despite the scrutiny of what goes on in the classroom, teaching remains very much an art form. Distance learning may depend even more on instructor charisma and style than the traditional classroom in which case instructor characteristics are important to examine in terms of their effect on interaction. It is axiomatic that the difference between a good teacher and a great teacher is the ability to motivate their students to learn (Main, 1992). Interpersonal communication skills are more critical when students are not physically present in the classroom. The technology of distance learning changes the dynamics of instruction. Beaudoin (1990) suggests distance education revolves around a learner-centered system with instructor skills focused on facilitating learning and organizing instructional resources.

Inserting technology in the instructional process requires greater attention to lesson design and instructional preparation. This factor needs to be more carefully examined and controlled in distance learning research. Miller (1989) argues that curriculum issues are more important than the delivery technology. Farr and Shaeffer (1993) provide a discussion on media selection variables for distance learning application.

Course variables include such things as the subject matter, student characteristics, instructional strategies and activities, media selection and instructor characteristics. Subject matter can be characterized in terms of type (cognitive, psychomotor, affective), depth or complexity (basic skills, advanced studies), application (practical, theoretical), level of proficiency (familiarity, mastery, automation) and domain (history, language skills, electronics, etc.). This listing is not

comprehensive. Each category is a compound variable and the dimensions provided are certainly not exhaustive. There may be other taxonomies of instructional techniques and subjects that may be more useful for hypotheses generation for a particular distance learning situation.

Student characteristics involve age, gender, motivation, prior knowledge and experience. An important consideration is whether enrollment is voluntary or required. Self-selection of distance learners and traditional classroom students contaminates many of the field studies reported in the literature. This may not be an important factor when distance learning is only used as an outreach for students unable or unwilling to attend residence courses. The relevant question is, do the students learn? How does interaction affect the learning for *these students*. It is when distance learning is being considered as an alternative for traditional classroom education and training that attention needs to be given to any differences between the comparison groups. To make generalizations about interaction effects for this use of distance learning, the differences in characteristics of students who select distance learning and those in the traditional class setting must be controlled.

<u>Summary</u>

The successful expansion of distance learning as an alternative to the traditional classroom is dependent upon the improvement of instructional design to approximate the richness of the interaction that occurs face-to-face. The technology for fully interactive distance learning is not the hurdle. The problem is how to elicit active participation by the learner whose interest in the particular subject and education, in general, is minimal at best. Interactivity seems to hold promise. We need to find the best techniques for achieving it in a cost effective manner. Hopefully, this taxonomy will serve as a useful tool in finding some answers.

<u>References</u>

Barron, D. D. (1978). "Faculty and Student Perceptions of Distance Education ," *Journal of Education for Library and Information Science*, Vol. 27, p. 257-271.

Beare, P. L. (1989). "The Comparative Effectiveness of Videotape, Audiotape, and Telelecture in Delivering Continuing Teacher Education," *The Journal of Distance Education*, Vol. 3, No. 2, 58-66.

Beaudoin, M. (1990). "The Instructor's Changing Role in Distance Education," *The American Journal of Distance Education*, Vol. 4, No. 2, p. 21-30.

Berlo, David K. (1960). *The Process of Communication--An Introduction to Theory and Practice*, Holt, Rinehart and Winston, Inc., New York, NY.

Boak, C. and D. Kirby (1989). "Teaching by Teleconference: What Goes On," *Proceedings of the 8th Annual Conference of the Canadian Association for the Study of Adult Education*, University of Ottawa, Cornwall, Ontario, Canada, p. 26-32.

Chu, Godwin C. and Wilbur Schramm. *Learning from Television*, (4th Ed.), National Association of Educational Broadcasters, Washington, D.C.

Clark, R. (1983). "Reconsidering Research on Learning from Media," *Review of Educational Research*, Vol. 53, p. 445-459.

Dale, Edgar (1978). *The Good Mind*, Phi Delta Kappa Educational Foundation, Fastback 105, Bloomington, IN.

Farr, C. W. and J. M. Shaeffer (1993). "Matching Media, Methods, and Objectives in Distance Education," *Educational Technology Journal*, July, p. 52-55.

Fulford, Catherine P. and Shuquiang Zhang (1993). "Perceptions of Interaction: The Critical Predictor in Distance Education," *The American Journal of Distance Education*, Vol. 7, No. 3.

Fuller, Buckminster (1966). "What I Have Learned, How Little I Know," *Saturday Review*, Vol. 49, No. 46, Nov. 12, p 6.

Grant, F. L.(1994). Personal conversation with the author regarding the feasibility of integration of corporate training and technical information services into virtual classroom and library resources at the Idaho National Energy Laboratory.

Haynes, Kathleen J. M. and Connie Dillon (1992). "Distance Education: Learning Outcomes, Interaction, and Attitude, *Journal of Education for Library and Information Science*, Vol. 33, p. 35-46.

Hefzallah, Ibrahim M. (1990). "The Learning Environment," Chap 2 of *The New Learning and Telecommunications Technologies*, edited by Ibrahim M. Hefzallah, Charles C. Thomas, Pub., Springfield, IL.

Hennings, Dorothy G. (1975). *Mastering Classroom Communication--What Interaction Analysis Tells the Teacher*, Goodyear Publishing Co., Pacific Palisades, CA.

Highet, Gilbert (1957). *The Art of Teaching*, Vintage Books, New York, NY.

Keene, D. S. and J. S. Cary (1990). "Effectiveness of Distance Education Approach to U.S. Army Reserve Component Training," *The American Journal of Distance Education*, Vol. 4, No. 2, p. 14-20.

Klinger, Toby H. and Michael R. Connet (1992). "Designing Distance Learning Courses for Critical Thinking," *T.H.E. Journal*, October.

Kozma, R. B. (1991). "Learning with Media," *Journal of Educational Research*, Vol. 61, p. 179-211.

Main, Robert G. (1992). *Integrating the Affective Domain in the Instructional Design Process*, U.S. Air Force Technical Report, Armstrong Laboratory, Brooks Air Force Base, TX.

Miller, J., W. C. McKenna and P. Ramsey (1993). "An Evaluation of Student Content Learning and Affective Perceptions of a Two-Way Interactive Video Learning Experience," *Educational Technology*, June, p. 51-55.

Moore, Michael, 1989. *Effects of Distance Learning: A Summary of the Literature*, OTA contractor report, U.S. Dept. of Commerce, National Technical Information Service (NTIS), Springfield, VA.

Ritchie, H. and T. J. Newby (1989). "Classroom Lecture/Discussion vs. Live Televised Instruction! A Comparison of Effects on Student Performance Attitude and Interaction, *The American Journal of Distance Education*, Vol. 3 No. 3, p. 36-45.

Rupinski, T. E. (1991). *Analysis of Video Teletraining Utilization, Effectiveness, and Acceptance*, CRM Research Memorandum 91-36, Center for Naval Analysis, Alexandria, VA.

Salomon, G. and R. E. Clark (1977). "Reexamining the Methodology of Research of Media and Technology in Education;" *Review of Educational Research*, Vol. 47, No. 1, p. 99-120.

Simpson, Henry, H. Lauren Pugh and Steven W. Parchman (1991). *Empirical Comparisons of Alternative Video Teletraining Technologies*, (NPRDC-TR-92-3), Navy Personnel Research and Development Center, San Diego, CA.

Stoloff, P. H. (1991). *Cost-effectiveness of U.S. Navy Video Teletraining System Alternatives*, CRM Research Memorandum 91-195, Center for Naval Analysis, Alexandria, VA.

Texas Higher Education Coordinating Board (1986). *Instructional Television: A Research Review and Status Report*, ERIC No. ED 276 422, Texas College and University System, Division of Universities and Research, Austin TX.

U.S. Office of Technology Assessment (1989). *Linking for Learning, A New Course for Education*, .. OTA-SET-4 30, U.S. Government Printing Office, Washington, D.C.

Van Haalen, T. and Miller, G. (1994). "Interactivity as a Predictor of Student Success in Satellite Learning Programs," *DEOSNEWS*, The Distance Education On-line Symposium, Vol. 4, No. 6, ACSDE@PSUVM.PSU.EDU.

Wagner, E. (1990). "Instructional Design and Development: Contingency Management for Distance Education," In *Contemporary Issues in American Distance Education*, Michael Moore (Ed.), Pergammon Press, p. 298-311.

Wetzel, C.D., Radke, P.H. and Stern, H.W. (1994). Instructional effectiveness of Video Media. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Weingand, D. E. (1984). "Telecommunications Delivery of Education: A Comparison with the Traditional Classroom," *Journal of Education for Library and Information Science*, Vol. 25, p. 3-12.

Zigerell, J. J. (1986). *A Guide to Telecourses and Their Uses.* ERIC No. ED 280 439, Coast Community College District, Cost Mesa, CA.

A NOVEL DESIGN CONCEPT FOR A SMALL, FORCE REFLECTING,
TWO DEGREE OF FREEDOM JOYSTICK

Augustus Morris, Jr.

Associate Professor

Department of Manufacturing Engineering

Central State University

Wilberforce, OH 45384

A NOVEL DESIGN CONCEPT FOR A SMALL, FORCE REFLECTING, TWO DEGREE OF FREEDOM JOYSTICK

Augustus Morris, Jr.
Associate Professor
Department of Manufacturing Engineering
Central State University

Abstract

A force reflection technology has been a source of research at the Armstrong laboratory for over a decade. It has been applied to areas ranging from simulation of complex G-fields in a static 1-G environment, to the development of force reflection joysticks for general scientific research in manual control, to looking at the specific application of force reflection joysticks to aid people with muscular spasticity.

It is desired to design smaller, force reflection joysticks aimed at specific, practical applications. However, it is also desired for these new designs to retain as much adaptability as the previous designs. Unfortunately, adaptability of these joysticks come at a cost not feasible for everyday applications.

This effort introduces the use of magnetic particle brakes as a possible actuator for smaller, force reflection joysticks aimed at specific, practical applications. The overall cost of such joysticks could be as little as one tenth the cost of a conventional force reflection joystick.

A NOVEL DESIGN CONCEPT FOR A SMALL, FORCE REFLECTING, TWO DEGREE OF FREEDOM JOYSTICK

Augustus Morris, Jr.

Introduction

A dual axis force reflecting joystick was constructed at the Armstrong Aerospace Medical Laboratory at Wright-Patterson Air Force Base, Ohio in 1991. This is the second generation of a similar device designed in the early 1980's [1]. At that time, the joystick was a single axis controller and the force reflection was accomplished through a pneumatic actuator.

It was discovered in the 1980's that by controlling the force reflection envelope felt by the human at the controller, an improvement in manual control efficiency was possible. For a compensatory tracking task with a sum of sines disturbance input, it was found that this first generation joystick allowed the human operator to behave as an optimal controller [2]. This prompted the Armstrong Laboratory to construct a smaller, portable two degree of freedom joystick interfaced to a microcomputer [3].

Since the construction of this device, research at the Armstrong Laboratory has been done to determine if such a controller could also be useful to aid individuals with muscular spasticity in

controlling their environment more efficiently [4]. A direct application of this technology would be the use of such a joystick to control a motorized wheelchair in order to give these individuals safe, independent transportation.

Research conducted last summer at the Armstrong Laboratory addressed the feasibility of reducing the size of the current force reflecting joystick even further. This design would be directly applied to its use in motorized wheelchairs. The study concluded, in particular, that DC torque motors would be the ideal actuator to use when high fidelity of the force restraint, simplicity in maintenance and control, and cost were the major factors. This type of design would be well suited for highly controlled experimentation, where, precise force reflection information is used to further basic research in this area.

However, this same design would be very impractical as a day to day assist device individuals with muscular spasticity. The cost for such a design would render the device unmarketable. What is needed is a second look at the design of a force reflection joystick which would still meet the needs of individuals with muscular spasticity, yet at an affordable cost. Such a joystick may not be as versatile as the design with DC torque motors. But, if the overall result is even sub-optimally achieved, the design would be a success.

With the advent of an inexpensive, force reflection joystick, two different type of studies would be possible. When carefully controlled, basic research is being conducted, the force reflection joystick designed at the Armstrong Laboratory has the versatility and precision for such studies. However, for studies of practical field applications of a force reflecting joystick, the more economical design could be employed.

Problem

The prime problem of this effort concerns the cost reduction of a force reflection joystick so that practical applications can be implemented. The current design using DC torque motors is a good design. However, it is a costly one. The advantages of the current design include wide variability of the force restraint, a wide variability in the spring restraint, dashpot restraint, and the mass restraint. The DC torque motors used as the actuators for the current design provide low maintenance, yet it is easy to control the torque of the motors. These characteristics of the joystick design using DC torque motors are very attractive when considering the individuals whom may be using this joystick.

As with any device, where there is human interaction, the human component interjects wide variability which must be accounted for. In most cases, a custom design is needed for optimal results between that individual, and the interacting device.

However, it is the high adaptability that gives the current design so much power. It also introduces high cost. The average servo motor, for applications such as this, with enough variability to accommodate most individuals would cost on the order of $2000.00 or more. Thus for a two degree of freedom force reflection joystick using DC torque motors, it is conceivable that the entire system could, run at a minimum cost of $5000.00. This is out of the price range for any practical application of this device. Therefore, a more economical design must be sought, which would contain most, if not all of the desirable characteristics as design using the DC torque motors.

Other actuation systems have been explored in the past. Such systems included hydraulic actuators and pneumatic actuators. In each of these actuator types, cost is still a major factor. Therefore, there could be no reduction in the overall cost of a force reflecting joystick using these devices.

Other possibilities included using a lower torque servomotor with a gear reducer in order to increase the torque employed at the joystick. Such a design has great appeal. However, when close analysis is done, backdrivability problems exists if the gear ratio is high. Even if a low gear ratio is used, and backdrivability does not become a problem, large time delays are introduced to the system just due to the apparent inertia felt at the stick. This still could be very undesirable.

The problem can be summarized as the attempt to design an ideal force reflection joystick at a cost that is practical for general use. This may not be possible. The characteristics of an ideal force reflection system must be able to produce a variable external force, a variable mass restraint, a variable dashpot restraint, and a variable spring restraint.

The current design using DC torque motors can approach these ideals, with some limitations. In either case, the current design is capable of producing all four characteristics. In order to design a more economical system, some of these characteristics from the ideal force reflection system may be sacrificed or severely compromised. However, the objective of this effort is to seek a reasonable design that should fulfill the major objective of allowing control of a motorized wheelchair for individuals with muscular spasticity.

The resulting design could be severely compromised from the ideal force reflection system. However, with the current design at the Armstrong Laboratory, a test bed is available for studies where the ideal conditions exist. Or, direct and practical applications of the new design would be a cost effective way of implementing these designs. It could be a cost effective way of producing a workable solution.

<u>Approach</u>

Ideally, results from studies of people with muscular spasticity would yield information regarding the appropriate ranges of each type of force restraint for the force reflection joystick. However, there is not presently enough data to provide such recommendations. However, in order to select the appropriate type and capacity of actuator, specifications on the force restraints are needed as a starting points.

Therefore, the following specifications were based purely on arbitrarily choosing maximum force restraints thought to be encountered during the actual use of the joystick. Below, are the specifications of the force reflecting joystick selected.

1. The joystick would actually be 4 inches in length

2. The maximum angular displacement of the joystick would be 30 degrees from vertical

3. The maximum mass restraint would emulate a 10 lbm at the end of the joystick

4. The maximum dashpot restraint would emulate a 10 lb force when the joystick has an angular velocity of 19 rad/s

5. The maximum spring restraint would emulate a 10 lb force when the angular displacement of the joystick is 30 degrees from vertical

In order to determine the torque requirement of the actuator used in the design, a worst case situation must be analyzed. This worst case situation was assumed to exist when the joystick emulates the condition where all of the restraints are at a maximum. The joystick oscillates between +/- 30 degrees at the rate of 19 rad/s. In addition, the position of the joystick at this instance is 30 degrees from vertical.

After taking into account this worst case scenario, the maximum capacity of the actuator was determined to be approximately 13 lb-ft.

Once the torque rating of the actuator is known, the major objective would consist of a search for actuators that meet the torque requirement along with consideration of weight and space minimization.

Results

If the design of the ideal force reflection joystick was the major objective, the use of direct drive, DC torque motors would remain the actuator of choice. These actuators are simple in design, can be easily controlled, and is characteristic of high torque/weight ratio. However, these actuators still tend to be relatively large for the size constraints placed on the joystick itself. In turn, this would lead to a large, heavy joystick for use on a motorized wheelchair.

In order to size the joystick to the task, one or two changes must occur. The first one requires across the board reductions in the specifications outlined in the Approach. When this is done, a smaller actuator may be selected

The second change would involve the elimination of some of the force restraints from the ideal force reflection joystick condition. The strategy in this event would be to exclude all forms of force restraints which would lead to a degradation of tracking performance. Studies from the past indicate that one of the most established guideline in joystick design is the concept of a minimal mass joystick. Because of the second-order characteristics of the mass restraint, large time delays would be introduced in the manual control feedback loop, making stability a prime concern.

In addition to the recommendation of keeping the mass of the stick at a minimal level, the same studies indicate that viscous damping should be minimized also. In fact, it is the spring restraint that is the major factor in sound joystick design. However, the original objective must be revisited. For individuals with muscular spasticity, it is the viscous damping that provides the major contribution leading to improved tracking performance from these individuals. Therefore, for this application, the dashpot restraint along with the spring restraint should yield very positive results for individuals with muscular spasticity.

After the decision to eliminate the mass restraint from the joystick design, the use of magnetic particle brakes as an actuator became extremely appealing. Despite that magnetic particle brakes only provide a dashpot restraint, there are several characteristics which stand out. Magnetic particle brakes, when compared to DC torque motors can supply an equivalent torque with a higher torque/weight ratio, while consuming less electrical power, at approximately 1/10th the retail cost.

A method of providing a spring restraint has not been addressed up to this point. In the attempt to keep the size and weight of the joystick down, fixed springs or rubber bands could be implemented in such a way as to allow for quick changes at

various levels of spring restraints. Another alternative would be to use a smaller DC torque motor in conjunction with the magnetic particle brake. The motor would only supply the spring restraint, while the brake would supply the dashpot restraint.

There could easily be a number of different configurations leading to an acceptable design. A particular design to settle on would highly depend on the ranking of the constraints. In this case, compact size and affordability were the highest considerations. It is obvious that any design void of DC torque motors, while maintaining a dashpot and spring restraint, would be highly considered.

Conclusions

The results of this effort suggest that any practical design of a force reflection joystick to aid individuals with muscular spasticity can not behave as an ideal force reflection joystick. The cost of such a general joystick would be too cost prohibitive. However, the need to further basic research in this area exists, and the use of a general, wide range, high fidelity, force reflection joystick is necessary to this end.

It appears that the force reflection technology at the Armstrong Laboratory has evolved to where the lab has a force reflection joystick capable of producing each type of force restraint. However, in the quest of adding validity to the technology,

practical applications must be developed to demonstrate feasibility.

The indications are clear that both basic and applied research need to continue in this area. However, the results of this effort indicate that the feasibility of using magnetic particle brakes as actuators should be fully explored. It is proposed that this effort continues through a SREP grant.

References

1. Repperger D.W., Frazier J.W., van Patten R.E. A Smart Stick Controller Based on a Static Equilibrium Model. Proceedings of the 1984 National Aerospace and Electronics Conference (NAECON), 810-817.

2. Morris A.,Jr., Human Operator Performance in a Regulatory Tracking Task with an Active versus a Passive Control Stick. Ph.D. Dissertation, Wright State University, 1988.

3. Repperger D.W., Scarborough E.L., Chelette T.L. Construction of a Dual Axis Force Reflection Stick and Test Station. Air Force Tech Report AL-TR-1992-0041, November 1991.

4. Trimble J., Repperger D.W., Phillips C.A. Parameter Determination for Development of Spastic Tolerant Stick Controllers. Proposal to the Veterans Administration, accepted Dec 27, 1991.

COMPUTER CALCULATION OF RATE CONSTANTS
FOR BIMOLECULAR REACTIONS

Mark A. Novotny
Scholar/Scientist
Supercomputer Computations Research Institute
Florida State University
Tallahassee, FL 32306-4052

COMPUTER CALCULATION OF RATE CONSTANTS
FOR BIMOLECULAR REACTIONS

Mark A. Novotny
Scholar/Scientist
Supercomputer Computations Research Institute
Florida State University
Tallahassee, FL 32306-4052

ABSTRACT

The objective of this research was to attempt to calculate the rate constants for bimolecular reactions by combining the modified Angle-Dependent Line of Normals model with semi-emperical or *ab initio* quantum calculations for potential energy surfaces. The bimolecular reactions under consideration are those which are currently being investigated experimentally at the Environics Division at Tyndall Air Force Base. In particular, we concentrated mainly on the reactions $IO + Cl \rightarrow I + ClO$ and $CF_3I + OH \rightarrow CF_3 + IOH$. We found that the semi-emperical package MOPAC was not sufficiently accurate to estimate the potential for these reactions, and have begun calculating potential energy surfaces using the *ab initio* package GAUSSIAN. We also detail the similarities and differences between normal transition state theory and the modified Angle-Dependent Line of Normals model, and outline some of their underlying assumptions. In addition, relaxed MOPAC structures are presented for some of the molecules which are currently being experimentally investigated.

COMPUTER CALCULATION OF RATE CONSTANTS
FOR BIMOLECULAR REACTIONS

Mark A. Novotny

1. Introduction

The goal of the two months of research at Tyndall Air Force Base was to initiate a research effort to calculate gas-phase rate constants for reactions of interest to the Environics Directorate (ED). In particular, the focus was on calculating rate constants for bimolecular reactions which are under investigation by the experimentalists working at the ED. The rate constant models used included standard transition-state theory (TST) and a modification of the Angle-Dependent Line of Normals (ADLN) model which was published in [1]. These methods require that one obtain, either from semi-emperical or *ab initio* calculations, the potential energy surface for the bimolecular reaction.

The rest of the report is organized as follows. In Sec. 2 a brief description of the modified ADLN model and its use as a calculational tool is given. Its assumptions and the justifications for these assumptions are stated. The ADLN model is then compared and contrasted with TST. In Sec. 3 a discussion of the successes and failures of using MOPAC and GAUSSIAN to calculate the required potentials for the ADLN model are given. Sec. 4 presents a brief account of the results to date. Finally, Sec. 5 contains a brief conclusion and an account of plans to continue the research.

2. The modified ADLN model and TST

The thermal rate constant is given by

$$k(T) = \left[\frac{1}{\pi\mu}\right]^{\frac{1}{2}}\left[\frac{2}{k_{\mathrm{B}}T}\right]^{\frac{3}{2}}\int_{\varepsilon_{\mathrm{th}}}^{\infty}S_r(\varepsilon)\varepsilon e^{-\varepsilon/k_{\mathrm{B}}T}\,d\varepsilon \tag{1}$$

where ε is the relative translational kinetic energy of the two colliding molecules, μ is the reduced mass of the colliding molecules and is given by $\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}$, k_{B} is Boltzmann's constant, T is the temperature, $S_r(\varepsilon)$ is the reaction cross section, and $\varepsilon_{\mathrm{th}}$ is the threshold energy for reaction.

In Simple Collision Theory (SCT) for two hard-spheres with collisional diameter d one has [2]

$$S_r(\varepsilon) = \pi d^2\left[1 - \frac{\varepsilon_{\mathrm{th}}}{\varepsilon}\right]. \tag{2}$$

However, for a hard sphere colliding with a diatomic molecule we know that the threshold energy $\varepsilon_{\mathrm{th}}$ depends on the orientation of the diatomic during the collision, as well as on the fact that only the translational energy which is directed *along the direction of normals* contributes to the energy available for the reaction. This modification allows one to write the rate constant for a collision between an atom and a diatomic (with electronic structure for $\varepsilon_{\mathrm{th}}$ but no other stucture) as [1]

$$k_{\mathrm{MADLN1}}(T) = 2\pi D^2\left[\frac{8k_{\mathrm{B}}T}{\pi\mu}\right]^{\frac{1}{2}}\int_0^{\frac{\pi}{2}}e^{-\varepsilon_{\mathrm{th}}(\phi)/k_{\mathrm{B}}T}\sin(\phi)\cos(\phi)d\phi \tag{3}$$

where ϕ is the activated-complex bond-angle for the colliding system. When $\phi=0$ the collision is a head-on collision. The subscript stands for the Modified Angle-Dependent Line of Normals model number 1 of Ref. [1]. See Fig. 1 for a detailed picture of such a collision. The form for the threshold energy is assumed to be

$$\varepsilon_{\mathrm{th}}(\phi) = \varepsilon_{\mathrm{th}}^0 + f(\phi) \tag{4}$$

where $\varepsilon_{\mathrm{th}}^0$ is the energy difference between the potential for the reactants seperated by an infinite distance and the potential maximum for a head-on collision, $\phi=0$, of the atom with the diatomic. Hence $f(0)=0$.

28-4

The rate constant in Eq. (3) is appropriate for bimolecular gas reactions of the type

$$A + BC \rightarrow AB + C .\qquad\qquad (5)$$

Further generalizations to diatomics with structure and to the collision between two diatomics were presented in Ref. [1]. Although such complications were considered in the calculation of rate constants, in this report only the simple rate calculation of Eq. (3) will be used. We can also use Eq. (3) to calculate collisions between more complicated molecules, particularly if we let the collision radius for the reaction, D, depend on ϕ and take $D(\phi)$ into the integral of Eq. (3).

It is beyond the scope of this report to present a comprehensive discussion of transition state theory (TST). For a historical development of TST, see Ref. [3]. However, a brief sketch of the relevant details can be considered to be

1)   For a given chemical reaction there exists a surface in phase space that separates reactants and products.

2)   On the reactant side, all configurations are equally probable, and are distributed with a Boltzmann distribution function.

3)   Once the dividing surface has been crossed, the products do not re-enter the reactant space. The set of states that comprise the dividing surface is known as the **transition state**.

4)   The transition state is popluated with a Boltzmann distribution.

TST is also called **activated complex theory**. Of course, neither the modified ADLN or TST are the correct way of calculating rate constants. The correct way to calculate a reaction rate constant is to solve the electronic time-dependent Schrödinger equation for all possible collisions of the nuclei, and to then use these with the probabilities associated with each type of collision. Of course, such a course of action is impractical. Consequently, the first simplifying assumption that can be made is the separation of the nuclear and the electronic motions. This is possible because the nuclear masses are so much greater than the electronic masses, and, therefore, the nuclei move much more slowly. Thus, it is a reasonable approximation to suppose that the electron distribution depends *only* on the instantaneous

positions of the nuclei, and not on their velocities. This is the *adiabatic* or *Born-Oppenheimer* approximation, and yields an *effective* electronic energy and potential surface for the collision under consideration. Thus we can calculate the reaction rate constant by solving the time-independent Schrödinger equation for a very large number of configurations of the nuclei, and so generate the complete potential-energy surface for the reaction. This surface can then be used to perform classical trajectory calculations of the nuclei for a wide variety of initial reactant states and suitably average the results to obtain $k(T)$.

One of the further assumptions that go into both the TST and the ADLN model is that $k(T)$ can be calculated using *only* the potential at the *continental divide* of the potential. This is the surface of the potential where the reactants are down-hill in one direction and the products are down-hill in the other direction. Both theories then also assume that once the reactant system crosses the continental divide the products are formed. In addition, both theories deal with the zero-point quantization of the vibrational energy of the transition state and the reactant state in the same fashion.

The difference between the modified ADLN model and the transition state theory is two-fold. 1) TST assumes that the transition state is populated with a Boltzmann distribution in energy, including the rotational and electronic degrees of freedom. In contrast, the modified ADLN model assumes that the potential is zero until the continental divide is reached, at which time the potential jumps to the continental divide. Thus the Boltzmann distribution is assumed before the continental divide is reached, but the Boltzmann distribution is *not* assumed at the continental divide. (In both theories the Boltzmann distribution is *not* assumed for the products.) 2) TST assumes that all the activated complexes that form will be near the *saddle point* (or the pass on the continental divide), and that the walls of the pass are sharp enough and symmetric enough that they can very well be approximated by a parabolic potential in the coordinates perpendicular to the reaction pathway. The modified ADLN model makes no such assumptions. Let us examine these two differences between the theories.

First, is the transition state populated with a Boltzmann distribution? This may be close to a religious discussion, so let me expound on my beliefs on this subject. I believe that this assumption is

valid for unimolecular reactions [4] and for reactive processes which take place while the system is undergoing interactions with its environment (which may be a solvent or a solid) where a Kramers-like introduction of a macroscopic friction enters [5]. However, in the case where the barrier crossing proceeds essentially like a *free flight* over the barrier, I believe that the validity of the Boltzmann distribution for the transition state depends on *whether or not the saddle point is extremely steep*, as well as *explicitly on the colliding particles*. Since the transition state is almost always extremely short lived, one cannot obtain the Boltzmann distribution from the rearrangements of the activated complexes, but must obtain them from the initial Boltzmann distribution of the reactants. For example, consider the reaction

$$HI + D \rightarrow DI + H . \tag{6}$$

Since the iodine atom is extremely large, the deuterium atom can approach from almost any angle, and if the transitional energy of the collision along the line of normals between D and I is large enough the DI molecule will be formed. Consequently, in this case one should expect that the idea of a symmetric saddle point with steep walls is questionable. Thus, this reaction should be driven more by the the collision probabilities of D and I, and not by a saddle point. Similar conclusions hold for other reactions which involve two heavy atoms, such as in the substitution reaction where one isotope of Cl substitutes for another in the HCl molecule. *Ab initio* calculations [6] then show that the saddle point is again not extremely shallow. This is shown in Fig. 2. Of course, we must compare the shallowness of the saddle point with the average energy of collision between the reactants. Since we assume a Maxwell distribution of the reactants when they are far apart, the average collision energy is given by

$$E_{collision} = 2k_B T \tag{7}$$

where $k_B = 8.62 x \, 10^{-5}$ eV/K. (The constant 2 is for the average energy of translation in the center-of-mass (cm) frame, while the factor would be 3 in the frame in which the gas is at rest.) At room temperature, this means that the average collision energy in the cm frame is about 0.05 eV, and in the troposphere it may be only 0.04 eV. As seen in Fig. 2, this means than any collision which occurs in the range $0 < \phi < 25$ degrees may go to products. Thus the concept of a clean sharp saddle point breaks down.

A case in which an activated complex most likely will not be populated by a Boltzmann distribution can be understood by considering a reaction involving a light-heavy-light diatomic, such as IO. If

we treat IO as a two-body classical rigid rotor (at room temperature the average quantum number J of rotation is about 33, so this approximation should be fairly good), then the ratio between the kinetic energy (KE) due to rotation of O to the total KE of rotation is about 0.9 and the ratio of the linear velocities due to rotation of O and I is about 8. Consequently, almost all the energy of rotation is in the light atom, and the initial impact between this diatomic and a non-interacting hard sphere will *almost always* be between the hard sphere and the O atom. This analysis suggests that there should be much more kinetic energy available to overcome a saddle-point barrier if the saddle point is crossed when the incoming atom strikes the O atom then when the incoming atom strikes the I atom. It also suggests that the distribution at the saddle point may be different depending on whether the activated complex has the incoming atom striking the O atom, or whether the activated complex is formed with the incoming atom striking the I atom (since the incoming atom will almost always strike the O atom before it bounces to strike the I atom). Thus the TST assumption that the saddle point is always populated with a Boltzmann distribution is not valid for many reactions.

3. Quantum Calculations for use in the ADLN model

In principle there is no difficulty in applying quantum mechanics to calculate from first principles the rate constants for bimolecular reactions. However, there is a difficulty in practice, as is pointed out in a book by Benson published in 1960 [7]:

> With the advent of modern quantum mechanics in 1926, there followed a great activity on the part of physicists and chemists in attempting to calculate the bond energies of molecules. The results have been crude almost to the point of complete discouragement. The difficulty can be readily understood, from the point of view of the kineticist, in terms of the sensitivity of rate constants to activation energy. At 300K an error of 1.4 kcal/mole in E_{act} represents a 10-fold error in the rate constant. But 1.4 kcal/mole is about 0.07 eV, less by at least a factor of 10 than the errors involved in most of the available quantum-mechanical solutions to the Schrödinger equation for molecular systems.

Although this was written in 1960, and the availability and speed of computers has changed dramatically since 1960, still the bottleneck for calculating rate constants is the same today as it was three and one-half decades ago.

We first decided to try to utilize the semi-emperical package MOPAC [8] to obtain the continental divide of the potential. This decision was motivated by the fact that this was the only quantum package available at Tyndall Air Force Base, and there is currently no connection to the outside world from the computers at Tyndall Air Force Base. Using MOPAC first involved porting the MOPAC code to a SPARCstation 10, and going into the MOPAC subroutines and writing code to calculate the continental divide of the potential. Our reasoning was that if MOPAC could give us an estimate for the shape of the potential $f(\phi)$ of Eq. (4), then at least we should be able to obtain ratios of rate constants for reactions such as IO+Cl→ClO+I and IO+Br→BrO+I. However, it was only after the work of porting and modifying the code to obtain the desired quantities was completed that it became clear that MOPAC was not up to such calculations. We used the PULAY [9] option, as well as the PRECISE option. However, the accuracy was such that we obtained *extremely different* transition states depending on whether we brought in the Br atom to hit IO or the I atom to hit BrO (with identical angles formed by Br-O-I). This is because the potential was extremely insensitive to the atom-atom distance between IO and BrO. Similar difficulties were observed even when we restricted ourselves to light atoms such as C, O, and H. The difficulty is that, as pointed out above, we require a sensitivity of about 0.01 eV to obtain rate constants, while the typical value of the semi-emperical potential (we used the AM1 potential) is about 1000 eV for most reasonably large molecules.

After the failure to obtain results with MOPAC, we decided that we would next try the *ab initio* quantum chemistry package GAUSSIAN [10]. The first difficulty was that with the normal basis sets in GAUSSIAN (such as STO-3G and 6-31G) the heaviest atom which can be calculated is Ar, while we are interested in reactions containing at least one iodine atom. Consequently, we settled on the basis set LANL1MB [11] which can be used up to the atom Bi. LANL1MB has the basis STO-3G on the first row atoms, and the Los Alamos Encapsulated Core Potential plus MBS on the atoms Na through Bi.

However, the main difficulty was that access to GAUSSIAN from Tyndall Air Force Base was impossible due to the lack of network capabilities between the computers at Tyndall and the CRAY Y-MP at Florida State Univerisity where it was possible to obtain access to GAUSSIAN. Access to the CRAY Y-MP was obtained at the Florida State University branch campus located about 15 miles from Tyndall AFB. Consequently, we have just begun to obtain results using GAUSSIAN. However, it is clear that GUASSIAN cannot directly calculate the quantity of interest in the modified ADLN model. Namely, it cannot easily calculate the continental divide of the relaxed potential between reactants and products for a particular incoming angle between the reactant molecules. It can, however, use the LST (Linear Synchronous Transit) option to locate a maximum in the potential along a path connecting two structures. It can also calculate a GRID of the potential surface *with the variables outside the grid frozen*. However, the ADLN model requires the grid variables to be *fixed* and all other variables associated with the activated complex to be *relaxed*. Thus we are currently writing shell scripts to perform the desired calculations using GAUSSIAN and the GRID and LST options together with an optimization procedure external to GAUSSIAN. Within GAUSSIAN, we will also utilize higher-order methods such as the coupled-cluster method to obtain more accurate *ab initio* values for the threshold energy and the potential at the continental divide. This project will continue provided that the SREP (Summer Research Extension Program) proposal is funded.

4. Results, Discussion, and the Experimental values

Even with the difficulties outlined in Sec. 3, there were some fruitful interactions between the two people involved in the computational effort (M. A. Novotny and Capt. F. L. Wiseman) and the experimentalists at Tyndall AFB (Dr. Ray Wells, Ms. Dale Williams, and Mr. Stewart Markgraf). In particular, it was possible to provide MOPAC relaxed structures using the AM1 potential for the reactants CF_3I, CF_3CFHCF_3, as well as molecules with the formula C_6OH_{12} and $C_6H_{12}O_3$. (Due to space restrictions, not all results are included in the tables for these molecules.) Reactions of each of these molecules with OH are currently being studied by the experimentalists at Tyndall AFB. The relaxed structures of three of these molecules are shown in the tables. All units of length in the tables are in

Angstroms, the angles are in degrees, and the charge is in units of the charge on the electron. In standard MOPAC language, the bond length is between the atom and atom NA, the bond angle is between NA and NB, and the torsion angle is between NA, NB, and NC.

5. Conclusions and Plans for Future Work

The major conclusion reached thus far is that the MOPAC package is not sufficient to obtain the potential continental divide between reactancts and products in bimolecular reactions. Thus we are forced to use *ab initio* calculations such as GAUSSIAN-92 or the University of Florida program ACES-2. However, once the required potential surfaces are obtained the program to calculate the rate constants using TST and the modified ADLN theories have been produced. If the proposal to the SREP program is accepted, then these calculations will continue. In particular, we should soon be able to calculate the rate constants from the modified ADLN models using the GAUSSIAN results for the potential.

In addition, we propose to examine the underlying assumptions by which the TST and ADLN theories differ. In particular, the assumption that the transition state is Boltzmann distributed will be examined using the GAUSSIAN potential and motion on the potential surface for various reactants.

There is an additional assumption in the ADLN model that is not completely justified. Namely that the potential is zero until the continental divide is reached. This means that there is *no* channelling of the incoming reactants toward the saddle point. This is opposed to the TST assumption that almost all incoming reactants are channelled toward the saddle point. It may be possible to relax the assumption of no channelling in the ADLN model, and to still not have to perform averages over all incoming trajectories of the reactants. This approach will be explored by constructing a tent-like potential between the continental divide and the potential when the reactants are far apart, and solving the classical trajectory problem on this tent-like potential.

However, the main thrust in the future is to complete the computer calculations of the rate constants for the reactions of interest at Tyndall AFB. The reactions to be studied are those listed in Sec. 4. The calculated rate constants will also be compared with published data, such as those compiled in Ref. [12].

REFERENCES

[1] F. L. Wiseman and A. G. Rice, *J. of Chemical Education* **70**, 914 (1993).

[2] I. W. M. Smith, *Kinetics and Dynamics of Elementary Gas Reactions*, Chap. 3, (Butterworths, London and Boston, 1980).

[3] K. J. Laidler and M. C. King, *J. Phys. Chem.*, **87**, 2657 (1983).

[4] J. H. Beynon and J. R. Gilbert, *Application of Transition State Theory to Unimolecular Reactions*, (Wiley, New York, 1984).

[5] *Activated Barrier Crossing, Applications in Physics, Chemistry, and Biology*, edited by G. R. Fleming and P. H Singapore, 1993).

[6] B. C. Garrett, D. G. Truhlar, A. F. Wagner, and T. H. Dunning, Jr. *J. Chem. Phys.*, **84**, 1730 (1983); as reported in Ref. [1].

[7] S. W. Benson, *The Foundations of Chemical Kinetics*, (McGraw-Hill, New York, 1960).

[8] J. J. P. Stewart, *MOPAC MANUAL* (Sixth Edition), Frank J. Seiler Research Laboratory, United States Air Force Academy CO 80840; October 1990.

[9] J. J. P. Stewart, P. Csaszar, and P. Pulay, *J. Comp. Chem.* **3**, 227, (1982).

[10] *Gaussian 92*, Revision A, M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1992.

[11] P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, **82** 270 (1985).

[12] R. Atkinson, *Kinetics and Mechanisms of the Gas-Phase Reactions of the Hydroxyl Radical with Organic Compounds*, *J. of Physical and Chemical Reference Data*, Monograph 1, (American Chemical Society, New York, 1989).

Table 1. Relaxed structure for CF_3I

ATOM#	ATOM	BOND LENGTH	BOND ANGLE	TORSION ANGLE	NA	NB	NC	Charge
1	C							0.275
2	I	2.18			1			0.164
3	F	1.37	114.1		1	2		-0.147
4	F	1.37	114.2	-120.2	1	2	3	-0.147
5	F	1.37	114.0	119.8	1	2	3	-0.147

Table 2. Relaxed structure for CF_3CFHCF_3

ATOM#	ATOM	BOND LENGTH	BOND ANGLE	TORSION ANGLE	NA	NB	NC	Charge
1	C							0.400
2	C	1.58			1			-0.013
3	C	1.58	111.1		2	1		0.400
4	F	1.37	114.5	-59.6	1	2	3	-0.136
5	F	1.37	111.8	-179.8	1	2	3	-0.144
6	F	1.37	112.6	61.6	1	2	3	-0.144
7	F	1.37	110.4	122.7	2	1	3	-0.101
8	H	1.13	107.4	-117.1	2	1	3	0.162
9	F	1.37	112.6	-60.8	3	2	1	-0.144
10	F	1.37	111.8	-179.3	3	2	1	-0.144
11	F	1.37	114.5	60.5	3	2	1	-0.136

ATOM#	ATOM	BOND LENGTH	BOND ANGLE	TORSION ANGLE	NA	NB	NC	Charge
		Table 3. Relaxed structure for $C_6H_{12}O_3$						
1	C							-0.218
2	C	1.49			1			0.302
3	O	1.37	112.7		2	1		-0.278
4	C	1.43	117.9	-178.8	3	2	1	-0.045
5	C	1.53	110.2	-81.1	4	3	2	-0.050
6	O	1.42	110.6	-173.0	5	4	3	-0.275
7	C	1.43	113.5	-82.5	6	5	4	-0.025
8	C	1.51	106.7	-178.5	7	6	5	-0.215
9	H	1.12	109.7	58.2	1	2	3	0.116
10	H	1.12	109.6	-61.0	1	2	3	0.116
11	H	1.12	109.4	178.7	1	2	3	0.116
12	O	1.23	128.7	-179.8	2	1	3	-0.356
13	H	1.12	110.3	41.8	4	3	2	0.100
14	H	1.12	103.5	160.0	4	3	2	0.119
15	H	1.12	110.0	63.3	5	4	3	0.091
16	H	1.12	110.4	-58.1	5	4	3	0.110
17	H	1.12	109.6	61.2	7	6	5	0.067
18	H	1.12	109.4	-58.1	7	6	5	0.071
19	H	1.12	110.4	59.7	8	7	6	0.087
20	H	1.12	109.3	179.8	8	7	6	0.080
21	H	1.12	110.5	-60.1	8	7	6	0.086

Figure Captions

1. The collision between an atom and structureless diatomic is shown. The impact parameter is b, and the velocity vector of the incoming atom A makes an angle ϕ on impact. The line of center velocity is $|\vec{v}|\cos(\phi)$, and it is only this component of velocity that can overcome the barrier energy ε_{th} necessary for the reaction to proceed.

2. The dependence of ε_{th} on the angle ϕ is shown for the reaction

$$^{37}Cl + H^{35}Cl \rightarrow H^{37}Cl + {}^{35}Cl$$
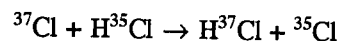
from Ref. [6]. The form shown is from *ab initio* calculations and is given by

$$f(\phi) = (9.2718x\,10^{-19})\phi^4 - (1.9551x\,10^{-19})\phi^2$$

where ϕ is in radians and the units for $f(\phi)$ are Joules/molecule. Note how shallow and asymmetric the saddle point is, which makes one believe that transition state theory may not be the best approximation to use to calculate $k(T)$, but that a modified ADLN model may be better.
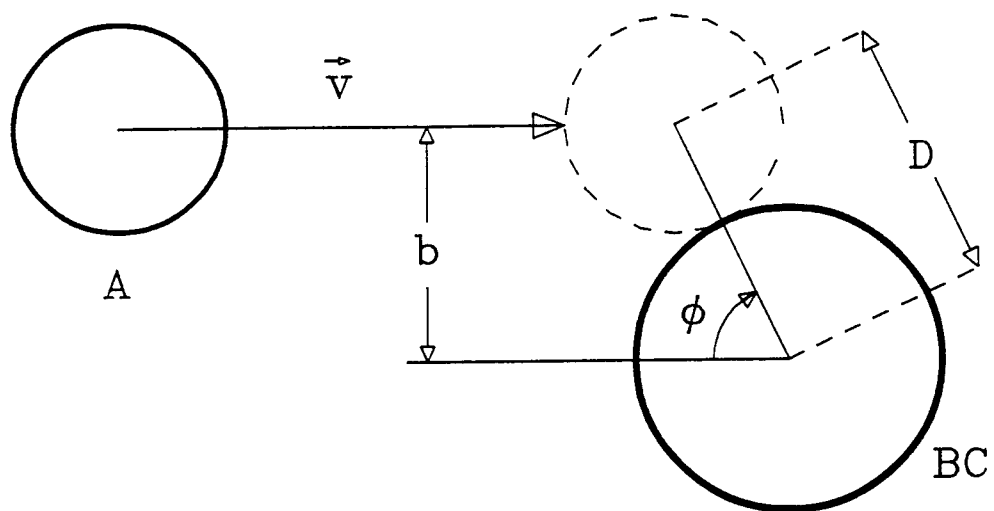
Figure 1.
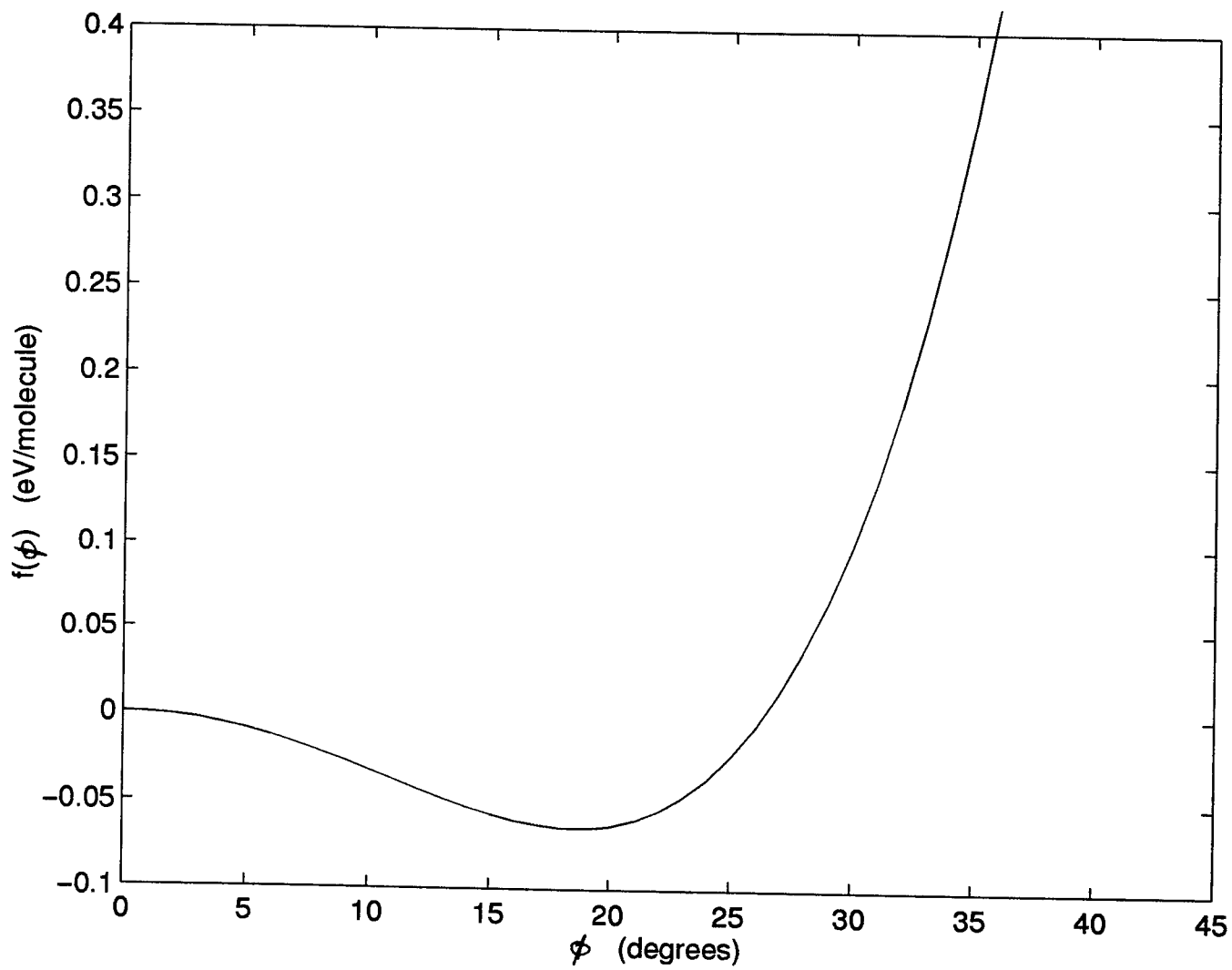
Figure 2.

A REVIEW OF PARAMETER SELECTION FOR PROCESSING CYLINDRICAL HEAD SCAN DATA

Joseph H. Nurre
Assistant Professor
Department of Electrical & Computer Engineering

Ohio University
Athens, Ohio 45701

A REVIEW OF PARAMETER SELECTION FOR PROCESSING CYLINDRICAL HEAD SCAN DATA

Joseph H. Nurre
Assistant Professor
Department of Electrical & Computer Engineering
Ohio University

Abstract

Full field surface data of cylindrically shaped objects, such as a human's head, can be acquired by rotating a triangulated laser and imaging system about the subject. The method of acquisition is imperfect and requires post processing of the data obtained. To eliminate rough or irregular surface data, a two-dimensional convolution can be used. To eliminate spikes or impulse noise in the data, morphological filters are available. Proper use of both operations requires knowing and understanding the operational parameters. This report presents an overview of these methods and discusses the optimal parameter settings that have been determined by experimentation.

A REVIEW OF PARAMETER SELECTION FOR PROCESSING
CYLINDRICAL HEAD SCAN DATA

Joseph H. Nurre

I. Introduction

Helmet systems have become an integral part of today's modern aircraft. Better performing helmets, in terms of protection and head mounted sensory devices, require exact three dimensional anthropometry. The latest research in this area employs systems, such as the Cyberware ECHODigitizer 4020 RGB/PS-D, which uses computer vision and laser scanning of human heads to achieve full field three dimensional range data. Helium-Neon Laser light is projected along a known trajectory and is located on the surface using two separate images. This image redundancy allows surfaces to be captured which might be obscured in a single image triangulation system. The scanning system rotates about the subject's head, acquiring over 100,000 three dimensional points on the surface in about 12 seconds. The resulting contour data is stored as an array, 512 x 256, of 16-bit integer radius values from the center of rotation. The resolution is 1.563mm in the vertical (longitudinal) direction while the resolution in the horizontal or latitudinal direction (0.7 degrees of arc) depends on the distance of the surface from the center of rotation. Typically, for the surface of the head, this resolution is about 1.5mm. The search for the laser illumination proceeds from a maximum allowable radial diameter to zero. As the subjects are scanned, stray external light sources, unexpected surface reflections and suspended dust particles result in spikes and rough surface data.

Post processing range data, to compensate for unwanted defects, has been an area of research for many years. The topic is usually included in the field of image processing and many techniques are available. One unique feature of the head scan data is its cylindrical coordinate system. This requires special software design techniques. The Computerized Anthropometric Research & Design Laboratory (CARD) at the U.S. Air Force's Armstrong

Laboratory has been developing software to manipulate and display the head scan data. The main program called INTEGRATE has a wide variety of functions, implemented in a modular programming fashion. Modules specifically designed to eliminate spikes and rough surface data will be addressed in this report.

This report is intended to be a tutorial on INTEGRATE's data correction modules. The modules for eliminating spikes or impulse noise will be presented in Section II. Section III will present the modules accessible for smoothing rough surface data. Theoretical concepts and experimental justifications are available in references presented in both Sections. Section IV will then present a discussion.

II. Removing Spikes from Surface Data

II.a Background

To remove impulse noise from head scan data, a popular image processing technique known as morphology is used. Mathematical Morphology is the probing of an image shape with a specified structuring element. The geometry of the structuring element implies certain geometric characteristics about the image being analyzed. Two morphological operations, erosion and dilation will be briefly introduced below. Further information on morphological techniques and the theory of Mathematical Morphology can be found in Dougherty[1].

To find the erosion of a signal by a structuring element, the element is placed below the signal and forced up to the highest elevation, which will touch but not cross the signal. The origin of the element becomes the new signal value. The process is repeated throughout the length of the original signal. Because of the cylindrical coordinate system, the number of points influenced by the structuring element in an erosion changes, depending on their distance from the zero radius axis. For an erosion, the initial position of the origin of the structuring element must be placed at a distance from the zero radius axis equal to the smallest radial value of the effected longitudinal scan, minus the radius of the sphere. The element was pushed up from this point with fewer points required for checking, as it moved.

Dilation is the dual operation to erosion. When dilating a signal with a structuring element, the signal becomes a path for the origin of the element. As the element translates its path, a new signal is created which is the minimum signal needed to bound the structuring element.

The geometry of the structuring element implies certain geometric characteristics about the image being analyzed. Choosing among the infinite number of structuring elements, in order to find an appropriate one for a given application, is an important part of Morphology. It has been shown that removing impulse noise from human head scan data can be reduced to removing outward pointing spikes with a diameter less than 4mm, with an omni-directional structuring element.[2] Such an element is a sphere shown in Figure 1. To remove spikes, a head scan is first eroded with the element.

Figure 1. A spherical structuring element satisfies the requirements for spike removal on the head scans.
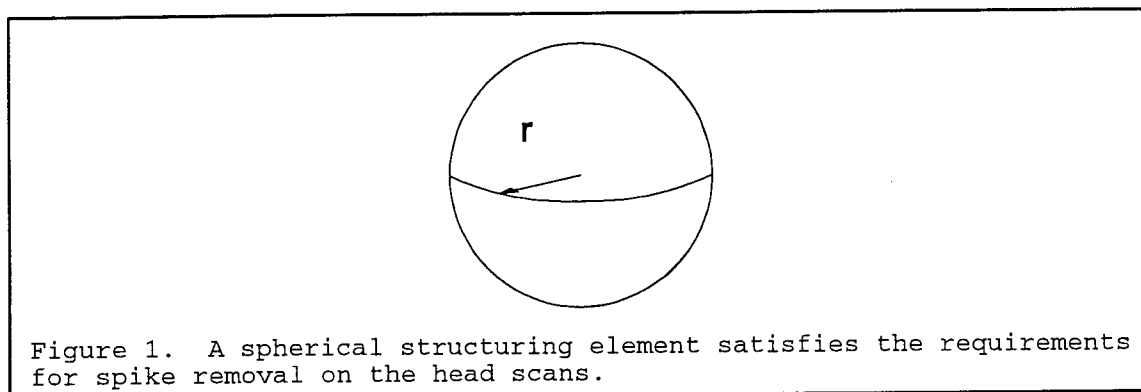
Failure of the element to fit into the spike results in its elimination. The head is then dilated back to its original size. An erosion operation followed by a dilation is commonly referred to as an opening.

II.b Examples

One of two goals are desirable when post processing head scan data with an opening. The first goal is to simply identify the spikes for further analysis. The second goal is to move the spike data to a location consistent with the surface. These goals are addressed in separate examples. Both examples will use the head scan presented in Figure 2. This head scan data,

one of the worst cases of spike corruption, is available as Cyberware file a017.3 from the HGU55P survey conducted by CARD.

To identify and remove spikes on the head scan, the following INTEGRATE commands could be executed:

```
CLOAD a017.3
TRIM 0 0 50 -55
COPY
ERODE 3
DILATE 3
ABSSUB 1 2
THRESHOLD 2 1 BELOW
AND 1 2
ABSSUB 1 2
```

The first steps of the example are to load the data file and trim away the top and bottom latitudes, as shown in Figure 2. This eliminates data

Figure 2. Cyberware file a017.3, from survey 55p, represents a head scan severely corrupted by impulse noise.
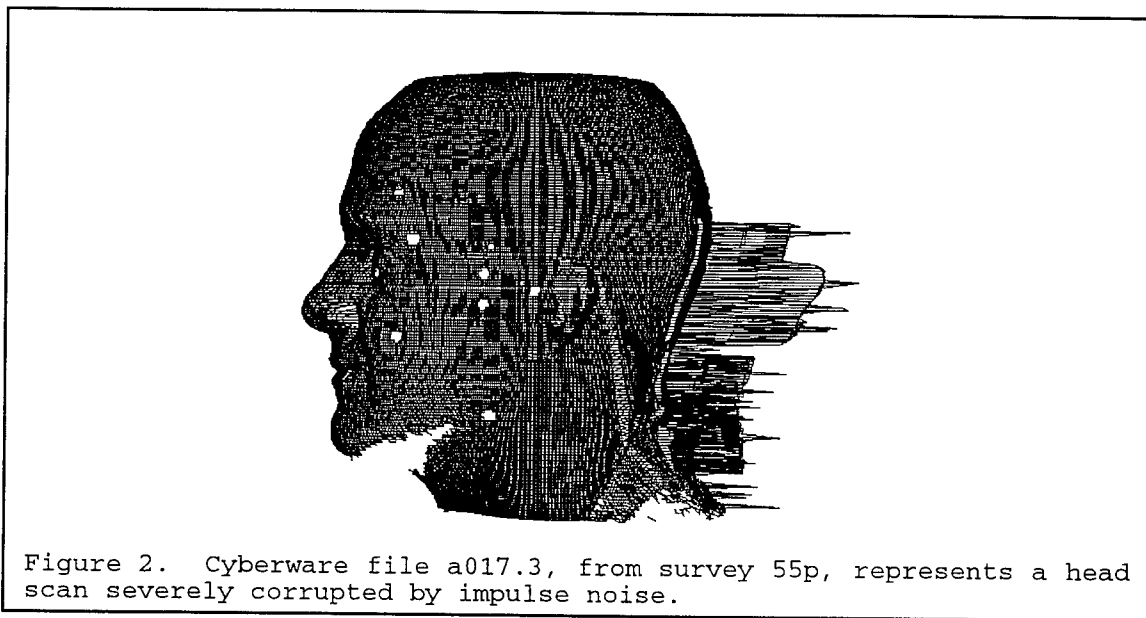
taken from the shoulder area and the top of the head positioning unit. It is important to discard this highly volatile data because a data point too close to the center of rotation can abort the erosion operation. As stated earlier, the element needs to be placed below the data points closest to the center. Points on the surface of the head have enough distance from center for the

element to fit.*

A copy of the head scan is made before the opening operation. An erosion with a sphere of size 3mm will eliminate approximately 98% of all spikes from the data. A subsequent dilation will only significantly change 1% of the data points that were not spikes.** Increasing or decreasing the sphere size will increase or decrease the number of spikes eliminated, while increasing or decreasing the number of good data points corrupted.[2] Figure 3 shows the average relationship between sphere size and affected
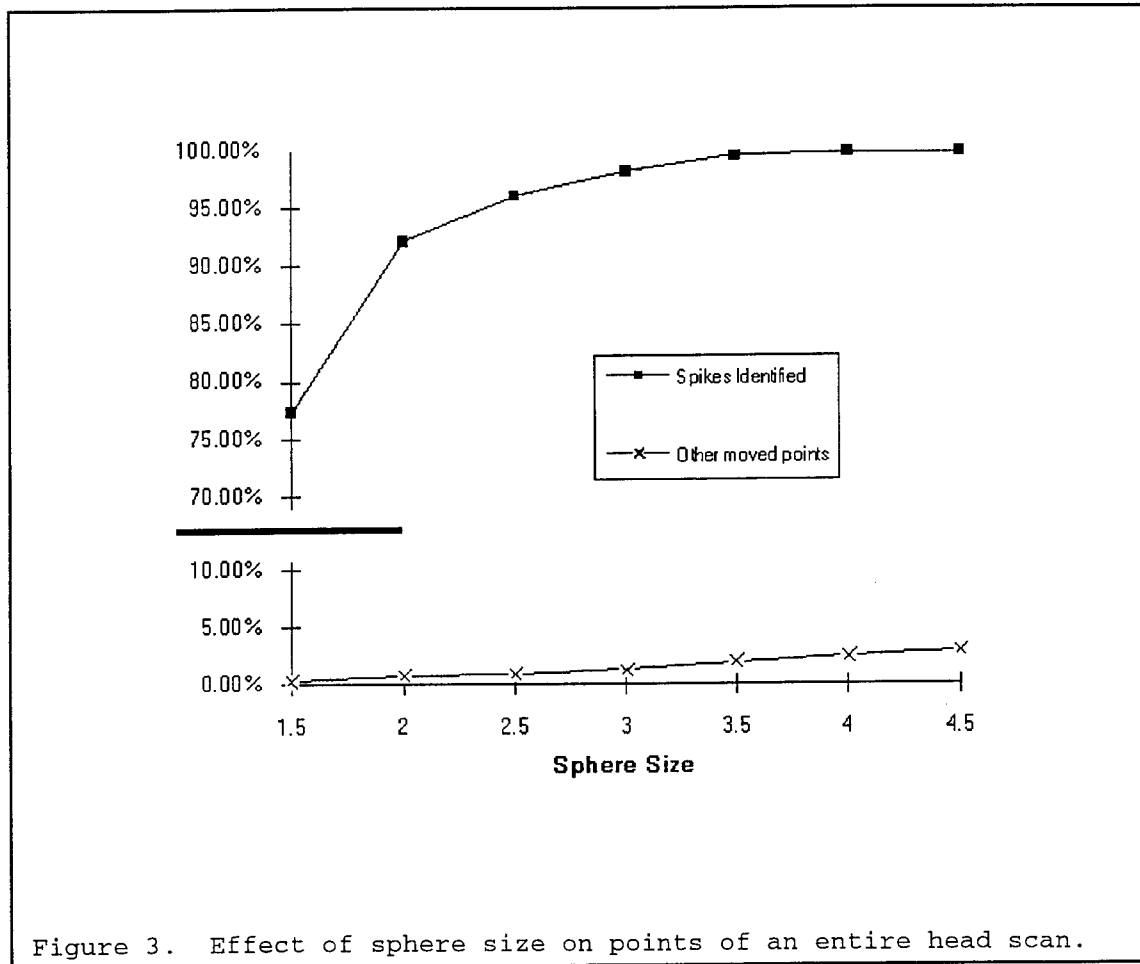
Figure 3. Effect of sphere size on points of an entire head scan.

---

* Surface locations that fail to be recorded by the Cyberware system result in data points with a radial value of zero. The erode module will ignore these points.

** A significant change is defined as movement of the points by more than 1mm.

points for data given in Nurre et.al.[2]

At this point, INTEGRATE has the original trimmed head scan in a data set referred to by the software as data set 1. The processed copy is in data set 2. An absolute subtract is performed on the two head scans with the result placed in data set 2. Points that were moved more than 1mm by the opening operation are set to zero using the threshold command on data set 2. Figure 4 shows the resulting data set of points identified as spikes.

Figure 4. Spike data from a017.3.
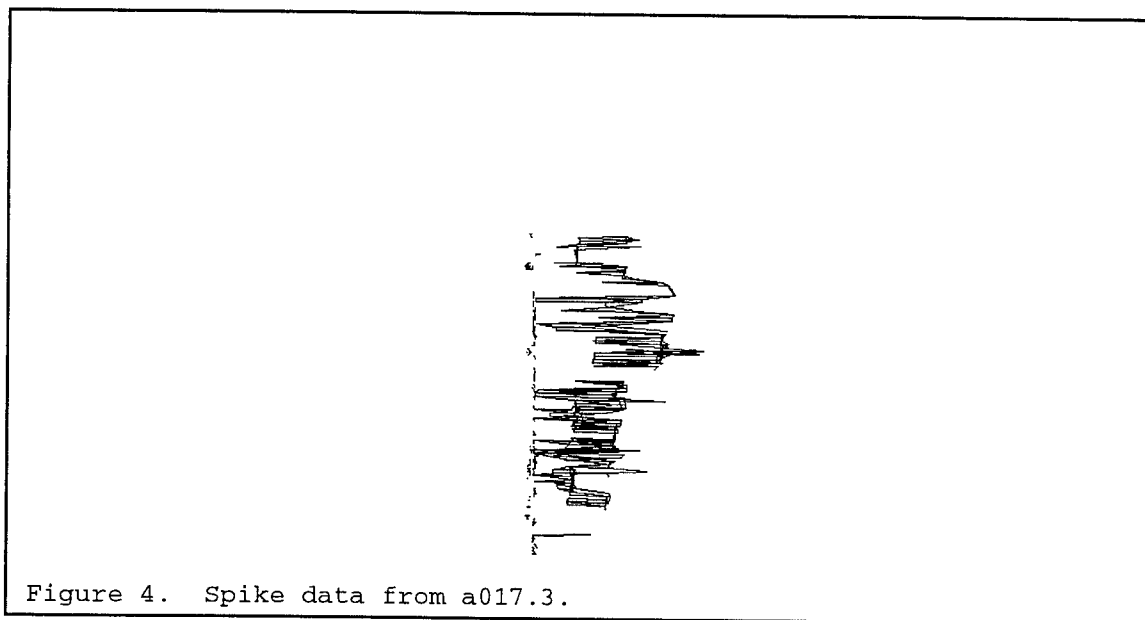
To remove the spikes from the original data an *AND* operation is performed. An *AND* operation creates a new data set with zeros where either or both of the operand sets had zeros. In this case, the new data set replaces data set 2 and any required non-zero values are taken from data set 1. Now data set 2 can be subtracted from data set 1 removing all identified spikes from the original data set. Figure 5 shows the final result. For the badly corrupted head in this example, increasing the sphere size will eliminate the remaining spikes but begin to effect the ear. Radial sizes between 3.0mm and 3.5mm represent a good trade off between identification and lost data.

For the example above, the spike points have been interpolated to new

Figure 5. Cyberware file A017.3 after being processed to remove spikes.
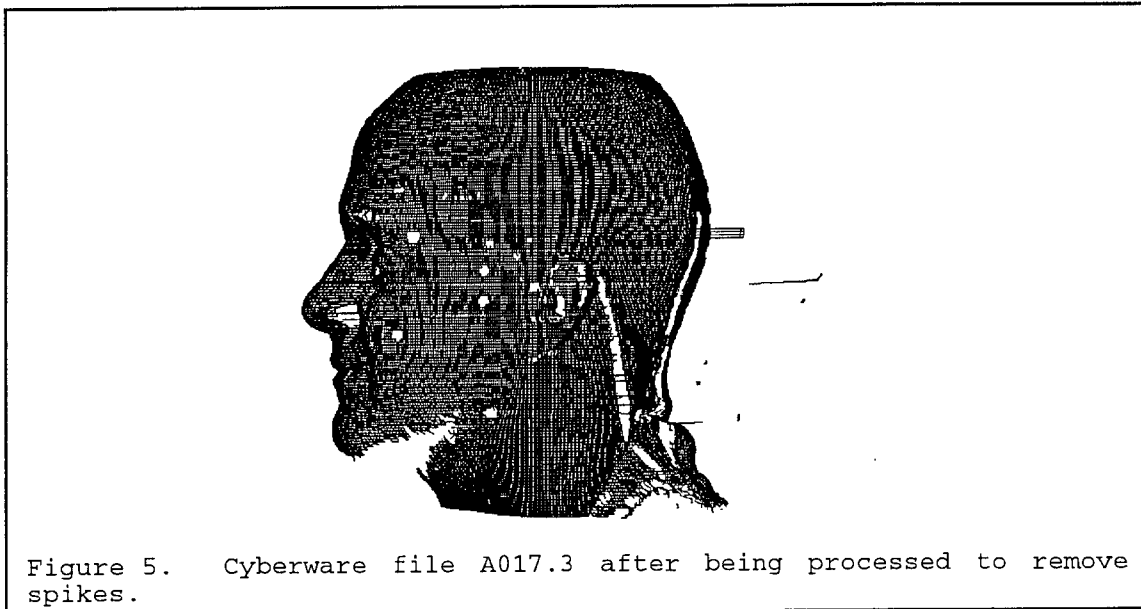
positions after the opening operation. An average of 86% of the interpolated points will be at their ideal location. To improve on this percentage, one must segment the head. Many of the points which are not correctly interpolated fall on regions of the head with complex surfaces such as the ears. Fortunately, a disproportionally large number of spikes are usually found on the back of the head. To bring points to their ideal location at a rate of 98%, the following INTEGRATE commands could be executed:

```
CLOAD a017.3
TRIM 428 -382 50 -55
ERODE 5.5
DILATE 5.5
```

In this example a simple TRIM operation was used to isolate the back of the head. Other methods are available in INTEGRATE to accomplish this task. The results are shown in Figure 6. The percentage of good data points that were moved is approximately 1.3%. Figure 7 shows the average relationship between sphere size and the affected points. From the figure it is apparent that the percentage of correctly interpolated points does not improve significantly beyond the 5.5mm size. The corruption of good points, however, continues to deteriorate in a linear fashion. An opening of 5.5 is a good trade off between accurate interpolation and corruption of good data points.

Figure 6. Cyberware file a017.3 after having spikes on the back of the
head interpolated to new locations.
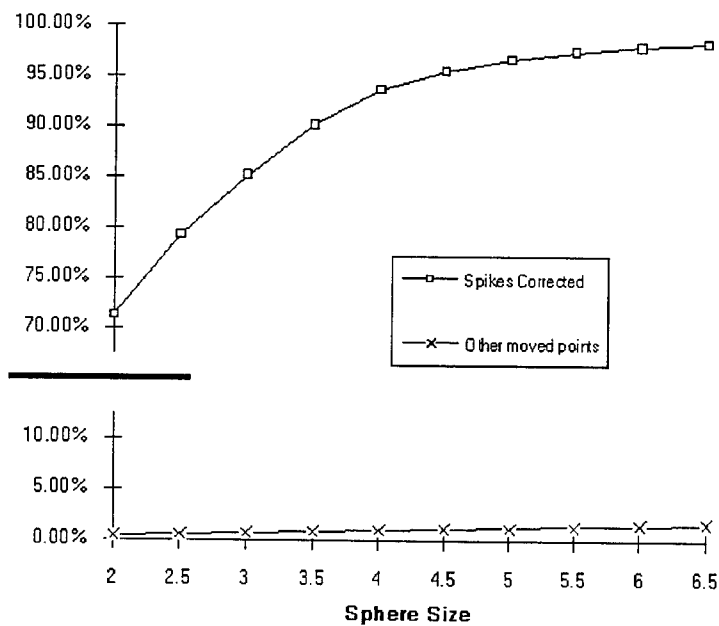
Figure 7. Effect of sphere size on points of the back of the head.

III. Smoothing Surface Data

III.a Background

Regularization is a theory developed early this century which has been used to solve ill-posed problems. An ill-posed problem may have multiple solutions. Regularization restricts the class of admissible solutions by introducing a "stabilizing function." An integral part of the regularization process is a parameter which controls the tradeoff between the "closeness" of the solution to the data and its degree of "smoothness" as measured by the stabilizing function. The mathematical expressions of regularization in one-dimension is given as follows:

$$E = \int_{\Omega} (f-d)^2 dx + \lambda \int_{\Omega} (\ f'\ )^2 dx \qquad (1)$$

The original data is d. The unknown fitting function, f, is determined by finding the minimum of E. The first term of the equation is a measure of the closeness of the solution function f, to the data d, while the second term is considered to be a stabilizer (a measurement of smoothness in terms of first derivative continuity). The compromise between these two terms is controlled by the regularization parameter λ.

Regularization does not give good results unless a method for determining an appropriate value for the regularization parameter exists. For practical purposes, it is desirable that such a method be data-driven and automatic. Generalized cross validation (GCV) derived from regularization offers an effective method for estimating the correct degree of regularization from the input data. Although the GCV is derived for determining the regularization parameter, it has been used extensively for determining the Gaussian filter parameter.[3] A comparison of the regularization filter with the Gaussian filter shows them to be nearly identical.

GCV is used to obtain an automatic algorithm for finding the optimal parameter of a discrete Gaussian filter from head scan data.[4] The

discrete Gaussian filter with the optimal parameter is then applied to obtain a smoothed estimate of head scan data.

To understand GCV, it is necessary to first define Mean Square Error (MSE). Assume the function g(t) is known and that n samples have been taken at t_i, i=1, ... ,n. If $g_{n,\lambda}(t)$ is an approximation to g(t) based on regularization, then the mean square error is defined as,

$$R(\lambda) = \frac{1}{n}\sum_{i=1}^{n} (g_{n,\lambda}(t_i) - g(t_i))^2. \qquad (2)$$

Equation (2) describes the MSE of the ideal data samples and the corresponding data samples from the regularized approximation function. Note that $g_{n,\lambda}(t_i)$ was determined by the uncorrupted data samples of g(t).

The ideal λ is defined as the λ which minimizes the true mean square error $R(\lambda)$, and is denoted by $\lambda*$. In practice, $\lambda*$ cannot be determined directly from $R(\lambda)$ since g(t) is not known a *priori*. Instead, what is usually available is:

$$y(t_i) = g(t_i) + e(t_i) \qquad (3)$$

where e(t) is additive white noise. A method for estimating $\lambda*$ given corrupted sample points is discussed below.

Let $g_{n,i}^{[k]}(t)$ be the approximating function constructed using all the corrupted data points, except the kth. The $g_{n,i}^{[k]}(t)$ can be used to predict the missing data point y_k, as a measure of the goodness of λ. Formally, let $g_{n,i}^{[k]}(t)$ be the function f(t) which minimizes:

$$E = \frac{1}{n}\sum_{\substack{j=1 \\ j \neq k}}^{n} (f(t_j) - y_j)^2 + \lambda\int_0^1 (f^{(m)}(u))^2 du \qquad (4)$$

The minimization of Equation (4) for a particular value of λ, leads to a convolution filter of size n, such that:

$$f(t_i) = \sum_{j=i-n}^{i+n} a_{ij} y_j \tag{5}$$

Now, the cross validation mean square error is defined as:

$$V_o(\lambda) = \frac{1}{n} \sum_{k=1}^{n} (g_{n,\lambda}^{[k]}(t_k) - y_k)^2 \tag{6}$$

The cross validation estimate of λ_o is the minimizer of $V_o(\lambda)$. The cross validation mean square error can be written in terms of the elements a_{ij} of the filter applied to y_i as:

$$V_o(\lambda) = \frac{1}{n} \sum_{k=1}^{n} \frac{\left[\sum_{j=1}^{n} a_{kj} y_j - y_k \right]^2}{(1 - a_{kk})^2} \tag{7}$$

The cross validation mean square error is a measure of the ability of $g_{n,\lambda}^{[k]}(t)$ to predict y_k, averaged over the values of k from 1 to n. A value of λ giving a minimal $V_o(\lambda)$ corresponds to an approximation of the samples which results in the best least squares prediction.

Generalization of the cross validation results in GCV. Details of this generalization and a discussion of the expansion into two-dimensions can be found in Fang and Nurre.[4] GCV can be thought of as finding the λ that minimizes a mean square error functional which is normalized to a filter described by λ. The GCV algorithm plots Equation (7) for the head scan of

interest and the user locates the minimum λ from the plot. It has been shown that the GCV produces a conservative estimate of λ*, the ideal filter size for uncorrupted data.

III.b Examples

In the first example, an entire head scan will be analyzed to determine the best filter size for smoothing rough surface data. The head scan to be used is in Cyberware file a177.3 from the HGU55P survey. To determine the best filter size to use, the following INTEGRATE commands could be executed:

```
CLOAD A177.3
TRIM 0 0 61 -50
{remove spikes if necessary}
INTERP LINEAR {or DO FILL w/ HISTOGRAM}
GCV DEFAULT
```

After loading and trimming the head scan, it must be checked for spikes. Impulse noise will invalidate the theoretical development of GCV. All spikes must be removed from the head scan data, if necessary. (A177.3 is already free of spikes.)

Data points in the head scan with zero radius value can be thought of as inward pointing spikes. These, also, must be removed. The DO FILL operation does the job quickly, but is not robust. In this example, a DO FILL was first attempted. A histogram of the data was taken and 2167 points were still set to zero. The head was reloaded into INTEGRATE and the INTERPOLATION command used. A new histogram showed no zeros.

With the head scan properly prepared, as shown in Figure 8, the GCV command is now executed. GCV will make a logarithmic plot of Equation (7). The parameters expected with the command are:

GCV {DEFAULT} or {START STEPS_DECADE TOTAL_STEPS}

START refers to the starting value of the plot. Because GCV plots logarithmically, the START value should always be specified as one times a power of ten. If such a value is not given, the largest power of ten, smaller than the specified value, is used. STEPS_DECADE refers to the number of points to be plotted per decade. TOTAL_STEPS specifies the total number of
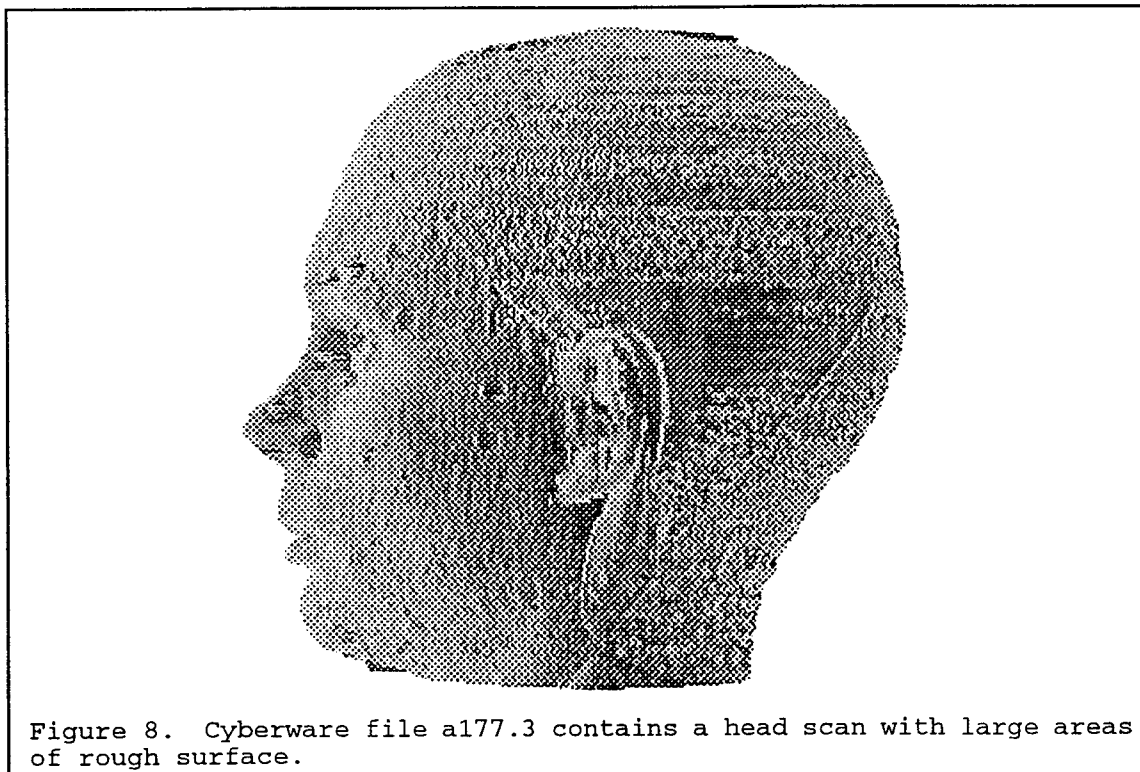
Figure 8. Cyberware file a177.3 contains a head scan with large areas of rough surface.

points in the plot. Using the command:

GCV DEFAULT

is equivalent to:

GCV 0.0001 10 100

Experience has shown that these values work best in most cases.

For the example, a default plot is made of the head scan. The plot, as it would appear in an INTEGRATE pop-up screen, is shown in Figure 9. The plot is on its side, with the logarithmic scale along the left most column. The generalized cross validation values are indicated with stars (*) from right to left with ascending value. The first of the three columns on the left contains the actual value calculated from Equation (7). Scanning the first column, it can be determined that the minimum plot value occurs at 0.0003. As stated in the previous section, this is a conservative estimate for the filter scale. In fact, the change in plot values is almost insignificant up to filter scale 0.2512.

To assist in choosing a reasonable scale value, the last two columns of

```
scale                                                    value    mean   Std Dev
-----   0.0                 1.0                 2.0      -----    ----   -------
        +-------------------+-------------------+
0.0001           *                                       :0.28274  0.00   0.0001
0.0001           *                                       :0.28143  0.00   0.0001
0.0002           *                                       :0.28160  0.00   0.0002
0.0002           *                                       :0.28122  0.00   0.0002
0.0003           *                                       :0.28087  0.00   0.0003
0.0003           *                                       :0.28107  0.00   0.0003
0.0004           *                                       :0.28109  0.00   0.0004
0.0005           *                                       :0.28114  0.00   0.0005
0.0006           *                                       :0.28112  0.00   0.0007
0.0008           *                                       :0.28105  0.00   0.0008
0.0010           *                                       :0.28114  0.00   0.0011
0.0013           *                                       :0.28101  0.00   0.0013
0.0016           *                                       :0.28109  0.00   0.0017
0.0020           *                                       :0.28110  0.00   0.0021
0.0025           *                                       :0.28108  0.00   0.0027
0.0032           *                                       :0.28113  0.00   0.0033
0.0040           *                                       :0.28116  0.00   0.0042
0.0050           *                                       :0.28120  0.00   0.0053
0.0063           *                                       :0.28126  0.00   0.0066
0.0079           *                                       :0.28135  0.00   0.0083
0.0100           *                                       :0.28144  0.00   0.0105
0.0126           *                                       :0.28156  0.00   0.0131
0.0158           *                                       :0.28172  0.00   0.0165
0.0200           *                                       :0.28192  0.00   0.0207
0.0251           *                                       :0.28220  0.00   0.0259
0.0316           *                                       :0.28253  0.00   0.0323
0.0398           *                                       :0.28298  0.00   0.0403
0.0501           *                                       :0.28355  0.00   0.0502
0.0631           *                                       :0.28431  0.00   0.0622
0.0794           *                                       :0.28531  0.00   0.0770
0.1000           *                                       :0.28664  0.00   0.0948
0.1259           *                                       :0.28844  0.00   0.1162
0.1585           *                                       :0.29088  0.00   0.1415
0.1995           *                                       :0.29422 -0.01   0.1711
0.2512           *                                       :0.29883 -0.01   0.2052
0.3162           *                                       :0.30523 -0.01   0.2438
0.3981          *                                        :0.31415 -0.01   0.2868
0.5012            *                                      :0.32660 -0.01   0.3338
0.6310            *                                      :0.34398 -0.02   0.3845
0.7943              *                                    :0.36817 -0.02   0.4384
1.0000               *                                   :0.40164 -0.03   0.4956
1.2589                 *                                 :0.44762 -0.03   0.5568
1.5849                   *                               :0.51019 -0.04   0.6230
1.9953                      *                            :0.59439 -0.05   0.6959
2.5119                         *                         :0.70625 -0.05   0.7776
3.1623                             *                     :0.85283 -0.07   0.8697
3.9811                                 *                 :1.04224 -0.08   0.9738
5.0119                                       *           :1.28387 -0.10   1.0909
6.3096                                             *     :1.58842 -0.12   1.2217
7.9433                                                 * :1.96786 -0.14   1.3666
```

Figure 9.   GCV plot for Cyberware file a177.3.

the plot contain the mean and standard deviation of the difference between the
original head scan and the filtered head scan. A mean close to zero is the
result of points on the filtered head scan moving to locations both inside and
outside of the original head scan by similar amounts.   If a normal
distribution is assumed, then 99% of all points are moved by less than 2.3
times the standard deviation. Based on these observations a filter scale of
0.5 is reasonable. This scale value approximates the end of the constant plot

values. Statistically, a scale of 0.5 will result in 99% of the data points being moved by less than 1mm.

The filter command to use would be:

FILTER DISCRETE .5 BOTH

Other filter options available in INTEGRATE are not appropriate with the current *GCV* algorithm. Figure 10 shows file A177.3, after being filtered. It should be noted that improvements in rough surfaces can best be seen with the INTEGRATE's *SURFACE ON* option, as shown in this figure.

Figure 10. Cyberware file a177.3 after being filtered with a discrete Gaussian of scale 0.5, in both latitudinal and longitudinal directions.
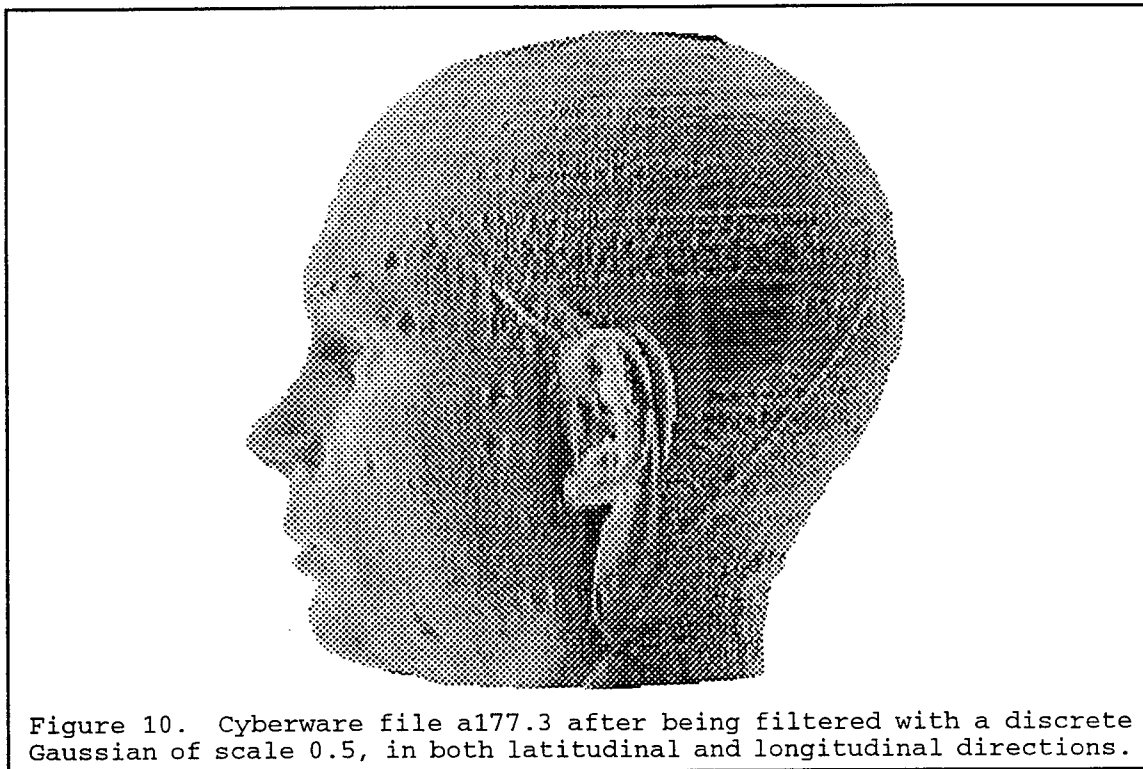
In the final example to be presented, the proper filter scale will be determined for a subset of head scan a177.3. It is the nature of generalized cross validation that intense noise in one segment of the head is averaged over the entire head, and therefore diluted. If surface roughness is not uniformly distributed throughout the head, than it is advantageous to segment the head. To determine the best filter size for a rough section of the head,

the following INTEGRATE commands could be executed:

```
CLOAD A177.3
TRIM 350 -272 147 -91
{remove spikes if necessary}
INTERP LINEAR {or DO FILL w/ HISTOGRAM}
GCV .01 20 60
```

A simple *TRIM* operation was used to isolate the section shown in Figure 11. Other methods are available in INTEGRATE to accomplish this task. In this example, the GCV plot is started at .01 and calculates 20 points per decade. Only 60 total points are to be plotted, resulting in the three decade plot shown in Figure 12. Examining this plot, a minimum value occurs for a filter scale of 0.5623. This being a conservative estimate, a scale of 0.9 was applied to the data. When applying a filter to a segment, the correct INTEGRATE command to use is:

FILTSEG   DISCRETE   .9   BOTH

The *FILTSEG* command is designed to keep the edge segment nearly stationary. This prevents the segment from becoming discontinuous with the rest of the head. Figure 11 shows the original segment while Figure 13 shows the filtered result.
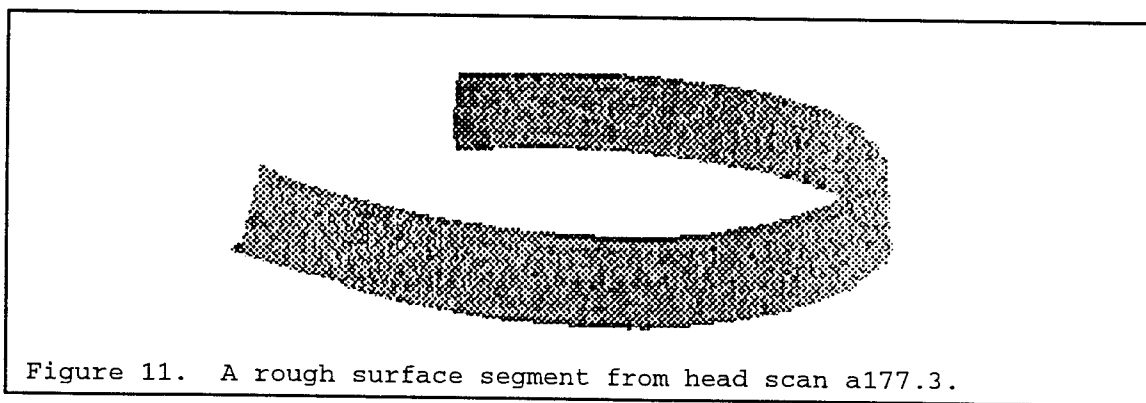
Figure 11.   A rough surface segment from head scan a177.3.

IV.   Conclusions

The examples presented above demonstrate the INTEGRATE commands now available for processing head scan data. Mathematical morphology operators

```
 scale                                               value     mean   Std Dev
 -----    0.0                    0.1              0.2  -----    ----   -------
         +--------------------------+-----------------------+
 0.0100                                               *    :0.18226   0.00   0.0084
 0.0112                                               *    :0.18223   0.00   0.0094
 0.0126                                               *    :0.18219   0.00   0.0106
 0.0141                                               *    :0.18214   0.00   0.0118
 0.0158                                               *    :0.18210   0.00   0.0133
 0.0178                                               *    :0.18205   0.00   0.0148
 0.0200                                               *    :0.18198   0.00   0.0166
 0.0224                                              *     :0.18191   0.00   0.0186
 0.0251                                             *      :0.18184   0.00   0.0208
 0.0282                                             *      :0.18176   0.00   0.0232
 0.0316                                            *       :0.18166   0.00   0.0259
 0.0355                                            *       :0.18156   0.00   0.0289
 0.0398                                            *       :0.18145   0.00   0.0323
 0.0447                                            *       :0.18132   0.00   0.0360
 0.0501                                            *       :0.18118   0.00   0.0401
 0.0562                                            *       :0.18103   0.00   0.0447
 0.0631                                            *       :0.18086   0.00   0.0497
 0.0708                                            *       :0.18068   0.00   0.0552
 0.0794                                            *       :0.18048   0.00   0.0613
 0.0891                                            *       :0.18026   0.00   0.0679
 0.1000                                            *       :0.18003   0.00   0.0752
 0.1122                                            *       :0.17977   0.00   0.0831
 0.1259                                            *       :0.17950   0.00   0.0917
 0.1413                                            *       :0.17922   0.00   0.1010
 0.1585                                           *        :0.17892   0.00   0.1110
 0.1778                                           *        :0.17860   0.00   0.1218
 0.1995                                           *        :0.17828   0.00   0.1333
 0.2239                                          *         :0.17795   0.00   0.1454
 0.2512                                          *         :0.17762   0.00   0.1583
 0.2818                                         *          :0.17730   0.00   0.1718
 0.3162                                        *           :0.17699   0.00   0.1858
 0.3548                                        *           :0.17672   0.00   0.2003
 0.3981                                        *           :0.17649   0.00   0.2152
 0.4467                                        *           :0.17631   0.00   0.2303
 0.5012                                        *           :0.17621   0.00   0.2455
 0.5623                                        *           :0.17620   0.00   0.2606
 0.6310                                        *           :0.17630   0.00   0.2755
 0.7079                                        *           :0.17652   0.00   0.2901
 0.7943                                        *           :0.17689   0.00   0.3042
 0.8913                                         *          :0.17741   0.00   0.3177
 1.0000                                         *          :0.17810   0.00   0.3305
 1.1220                                         *          :0.17897   0.00   0.3425
 1.2589                                         *          :0.18002   0.00   0.3536
 1.4125                                         *          :0.18125  -0.01   0.3639
 1.5849                                          *         :0.18266  -0.01   0.3734
 1.7783                                          *         :0.18424  -0.01   0.3821
 1.9953                                          *         :0.18601  -0.01   0.3901
 2.2387                                           *        :0.18794  -0.01   0.3975
 2.5119                                           *        :0.19006  -0.01   0.4043
 2.8184                                            *       :0.19236  -0.01   0.4107
 3.1623                                            *       :0.19488  -0.01   0.4168
 3.5481                                             *      :0.19761  -0.01   0.4226
 3.9811                                              *     :0.20061  -0.02   0.4283
 4.4668                                              *     :0.20389  -0.02   0.4340
 5.0119                                               *    :0.20752  -0.02   0.4397
 5.6234                                                *   :0.21154  -0.03   0.4455
 6.3096                                                 *  :0.21602  -0.03   0.4515
 7.0795                                                  * :0.22105  -0.04   0.4577
 7.9433                                                   *:0.22672  -0.04   0.4643
 8.9125                                                 *  :0.23314  -0.05   0.4712
```

Figure 12.   GCV plot, as specified by parameters given in the text, for a segment of Cyberware file a177.3.

were used to remove spikes.   Filtering for smoothing rough surface data is controlled by generalized cross validation.   Parameters for using the

Figure 13. The segment from Figure 11, after filtering with a 0.9 discrete Gaussian filter.
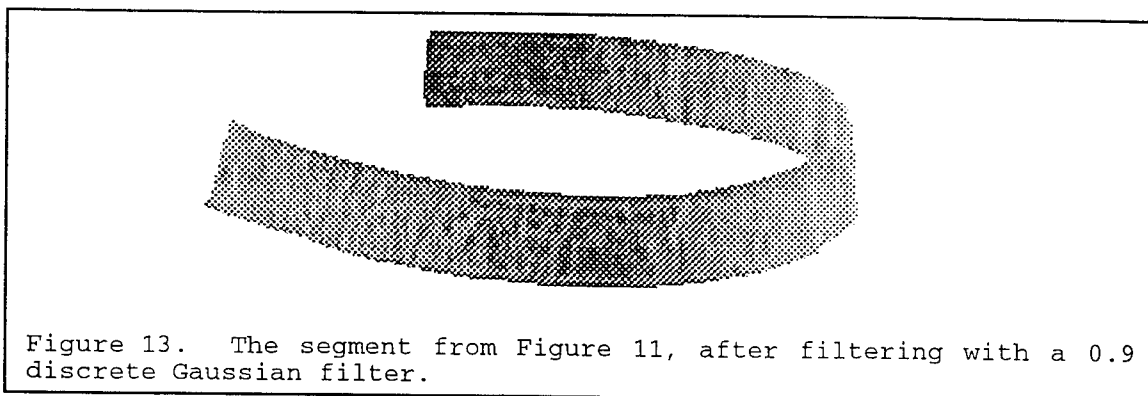
operators have been given to help quantify expected results. References were presented for users interested in the theoretical development. The INTEGRATE software continues to be at the leading edge of head scan analysis.

References

1. E.R. Dougherty, *An Introduction to Morphological Image Processing*, SPIE Optical Engineering Press, Bellingham, Washington, 1992.

2. J.H. Nurre, J.J. Whitestone, J.W. Hoffmeister and D.B. Burnsides, "Removing Impulse Noise from Human Head Scan Data", USAF Technical Report, In preparation.

3. D. Geiger and T. Poggio, "An optimal scale for edge detection", MIT Artificial Intelligence Laboratory Memo No.1078., 1988.

4. H. Fang and J.H. Nurre, "Optimal Estimation of Head Scan Data with Generalized Cross Validation", USAF Technical Report, Submitted, July, 1994.

IMPROVING THE UNITED STATES AIR FORCE ENVIRONMENTAL TECHNOLOGY DEVELOPMENT PLANNING

Edward L. Parkinson
Associate Professor
Department of Engineering Management

The University of Tennessee Space Institute
UTSI Blvd.
Tullahoma, Tn 37388

IMPROVING THE UNITED STATES AIR FORCE ENVIRONMENTAL TECHNOLOGY DEVELOPMENT PLANNING

Edward L. Parkinson
Associate Professor
Department of Engineering Management
The University of Tennessee Space Institute

Abstract

This paper presents salient points resulting from a summer faculty research study and technical report on technology development planning that emphasized the environmental area of "sludge". Improvement considerations are recommended in areas of organization, and strategic policies for long-range technology development and prioritization. These four recommendations should be applicable to all government organizations, development centers and industry dealing directly or indirectly with technology developments to meet environmental needs and objectives:

(1). Improve the "INTEGRATION and TEAMING" of functional areas.

(2). Adopt "RETURN-ON-INVESTMENT" (R.O.I.) as a primary component of the prioritization scheme for supporting research and development.

(3). Adopt the long-term thrust area of "TOTAL PROCESS CONTROL" for waste treatment and disposal as the ultimate reuse/recycle technology development planning objective.

(4). Adopt the planning objective of "COST COMPETITIVENESS" from the perspective of industry and government for use in decision processes in technology development including benchmarking and technology transition.

IMPROVING THE UNITED STATES AIR FORCE

ENVIRONMENTAL TECHNOLOGY DEVELOPMENT PLANNING

Edward L. Parkinson

INTRODUCTION

This paper is a result of a summer faculty research study sponsored by the Air Force Office of Scientific Research and the Air Force Armstrong Laboratory. This study reviewed technology development efforts related primarily to environmental sludge.

The study used generic technology development concepts for assessment and tailoring to the broad area of environmental wastes with a focus on environmental sludge. Aspects of the product life-cycle were included in assessment; such as, identification of needs and market analysis, trade-offs with other functional areas, technology transfer/transition, benchmarking and life-cycle cost considerations. This paper presents four recommendations resulting from the study consideration in technology development planning. The first relates to organizational improvement. The second is a strategy and approach for prioritization of resources in technology development. The third recommendation is a focus for long-term technology development objective. The fourth is an overall technology development strategy. The recommendations are discussed as follows:

(1).   Improve the "INTEGRATION and TEAMING" of functional areas.

(2).   Adopt "RETURN-ON-INVESTMENT" (R.O.I.) as a primary component of the prioritization scheme for supporting research and development.

(3).   Adopt the long-term thrust area of "TOTAL PROCESS CONTROL" for waste treatment and disposal as the ultimate reuse/recycle technology development planning objective.

(4).   Adopt the planning objective of "COST COMPETITIVENESS" from the perspective of industry and government for use in decision processes in technology development including benchmarking and technology transition.

"INTEGRATION AND TEAMING"

Improved internal/external organizational and functional area integration and teaming points to environmental technology development opportunities. Most entities whether at the federal, state or local levels organize to handle the environmental challenges into organizational areas of "remediation," "compliance" or "prevention." Initial emphasis has been in the clean-up of past environmental problems. This area is referred to as remediation. Another focus is on ensuring compliance with current and future regulations and requirements by manufacturing industry, operational and maintenance organizations, service organizations, and communities in general. The third organizational focus attempts to avoid or prevent the hazards and costs associated with existing processes and waste streams.

In the USAF, technology development for remediation and compliance are performed by the Air Force Armstrong Laboratory, Environics Directorate. As indicated in their mission statement, the mission of this directorate is to develop new technologies, processes, or models to reduce the cost in the cleaning up of past hazardous waste sites, while assuring cost-effective compliance with increasingly stringent environmental laws and regulations [1].

In addition there are also many other organizations that focus on these three areas of environmental technology development. The five Air Logistics Centers have organizations that deal with each of the three areas to meet world-wide logistics support requirements. Also, the five System Centers deal with each of the three areas for weapon system research, development and acquisition. In addition, the Aeronautical Systems Center Acquisition Environmental Management (ASC/EM) also has Air Force responsibility for the eleven Air Force owned and contractor operated facilities.

A breakout or remediation and compliance has merit for organizations charged with the actual management of the clean-up of existing problems and organizations charged with the reporting of compliance to the changing regulatory requirements. However, for organizations whose mission is for basic and applied technology and systems development, these organizational distinctions can present significant loss of potential efficiency and effectiveness in planning and implementation.

Although the degree of overlap between organizations is not known, figure 1 shows the perceived overlap for discussion purposes. The primary reasons for these overlaps stem from multi-use technologies, similar market opportunities for the basic and applied technologies being developed, and for similar functional area support needed during needs analysis and the phases of research and development (R&D).

Multi-Use Technologies (Remediation, Compliance and Prevention)

Review of documentation developed from the Department of Energy and the Environmental Protection Agency [2] [3], shows that over half of the technologies available have application

REMEDIATION

PREVENTION

COMPLIANCE

Figure 1. Conservative Estimate of Functional Area Overlaps for Technology Development.
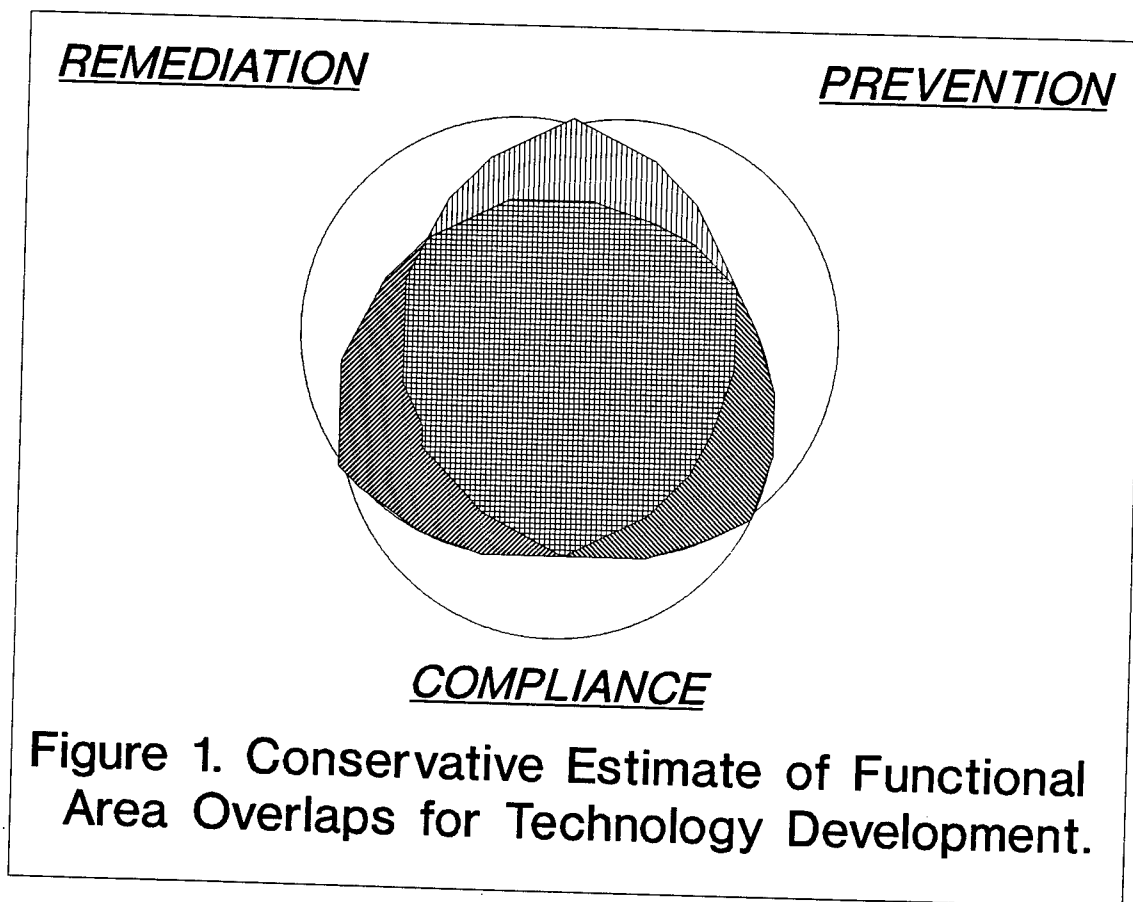
to wastes from manufacturing processes, as-well-as, soil treatments. This alone points to a strong basis for teaming and integration of remediation, compliance and prevention organizational and functional areas for planning the development of technology for environmental considerations. Regarding the focussed area of sludge, the data indicates that a large majority of these technologies are used for sludge treatment and disposal. Some of these technologies are categorized in figure 2.

Figure 2. Example of Sludge Treatment and Disposal Technology Categories.

Inside the figure:

CHEMICAL
BIOLOGICAL   AEROBIC
ELECTRO-KINETIC
DEWATERING
INCINERATION
ENCAPSULATION
VITRIFICATION
SOLVENT EXTRACTION
WASTE-TO-ENERGY
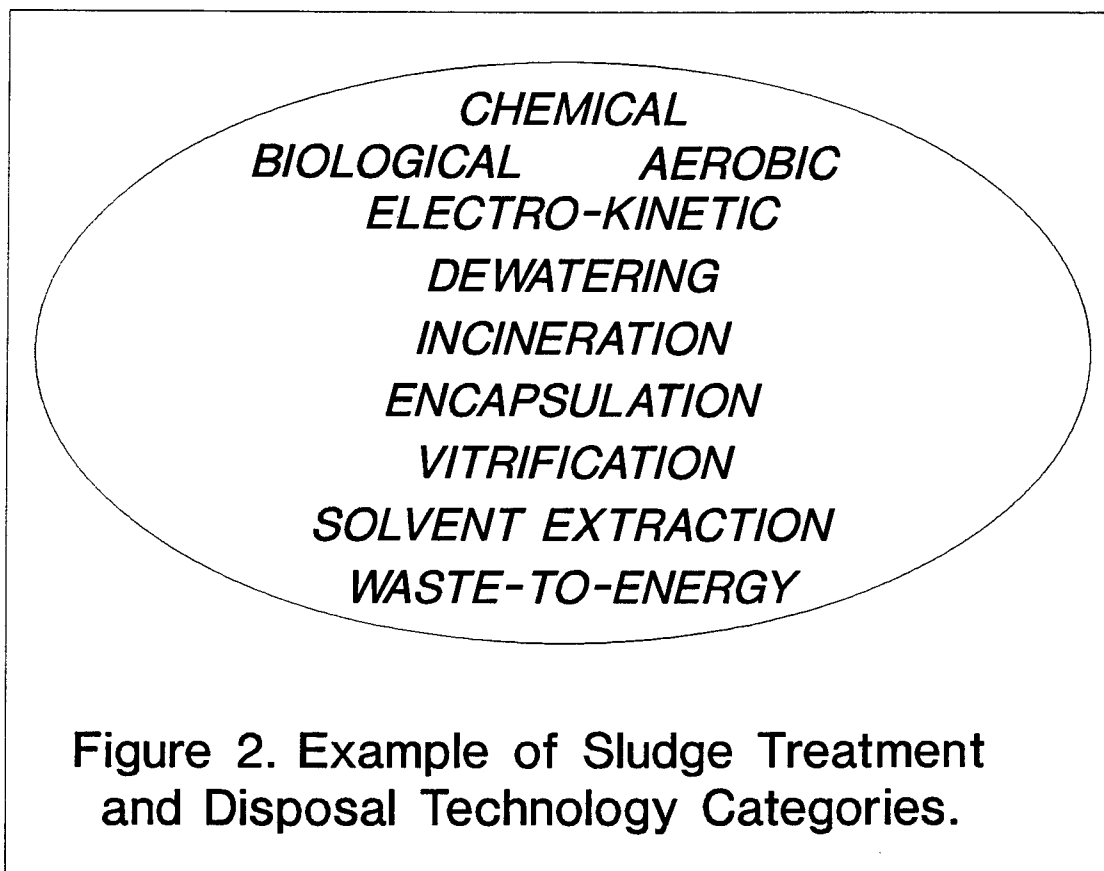
Increased Market

In general, one plans technology development to meet a particular market share of the total market potential for the technology. When it comes to Air Force development organizations, the direct market may be all the Air Force bases. However, this direct market potential increases when the development has direct application to other D.o.D. components and to other federal, state, local and industrial uses.

When one realizes that development of a basic technology can also be used directly or have spin-off applications to meet other environmental needs, this could also be included in strategy considerations for cost/benefit analysis and in organization integration and teaming.

30-7

When one can either use directly or tailor a best basic or applied technology development for various markets, the cost and benefits for the larger markets can become part of the technology planning process. In general, the larger market provides planning flexibility through some combination of either a basis for increased up-front investment in technology development or increases the return-on-investment.

In the environmental area, the markets for remediation, compliance and prevention will bring about significant increases in opportunity for technology development considerations. The literature indicates that the market for remediation is in the order of 75 billion by the year 2005 in the U.S. alone. This can be considered a large market with spin off's for technology development in the international market. However, it has been estimated that the U.S. corporations spend 115 billion annually in compliance with environmental regulations.[4]

The market potential for the Department of Defense alone is large for all three areas. Primary and support systems and facilities for air, land or sea missions must comply with environmental requirements. Depot maintenance and specialized repair operations are also a significant source of environmental wastes. Regarding sludge, many of the environmental wastes are either in the form of sludge or result in sludge through cleanup or disposal processes.

However, there is another market area, indirect source, where long-range technology planning and integration can have a significant impact and bring about positive contribution to effectiveness and cost. This indirect market source is in the development, manufacture, acquisition and procurement of weapons systems and follow-on logistics support. Although these environmental problems may not be physically within the D.o.D. for compliance, appropriate D.o.D. personnel can have a positive influence on environmental costs for manufacturing and

consequently D.o.D. organizational budgets. For example, the USAF has been spending 18-20 billion dollars per year for these procurements. The processes in industrial facilities used for manufacturing these systems and equipment produce wastewater and hazardous waste, whose treatments produce sludge. In many cases, the primary market for these manufacturers are the government logistics support needs of the Air Force and other D.o.D agencies. In these cases, any technology transfer of capability that improves the cost baseline, should result in downstream savings in the government budgets. How much savings would this be? It is probably not known what amount of the cost for procurements go for functional areas, such as; reliability, maintainability, human factors considerations and safety, let alone environmental. However, if the cost is only 1/2 of 1 percent for the environmental portion of the primary industrial process, cleanup and waste disposal, this market opportunity would be *"100 million dollars per year."* Based on the D.o.D. projection of about 45 billion dollars for the 1995 budget, this would provide a direct budget environmental impact opportunity for potential savings of *"225 million dollars per year."* Also, note that with added future mandated requirements this figure could easily be increased by an order of magnitude. In fact, a recent review by the United States Air Force Construction Cost Management Directorate suggested that the environmental functional area portion of costs could range from 5-10 percent of manufacturing costs. If this is accurate, the current D.o.D. market opportunity could be as high as *"4.5 billion dollars per year."* The main point here is that this indirect market represents a significant opportunity for return-on-investment consideration and can be use to support, justify and prioritize technology development efforts.

The coordination of teaming strategies with other D.o.D agencies, could potentially target a tremendous return-on-investment. As a note, from a national manufacturing and support basis, these cost saving opportunities would only be the tip of the iceberg when technology transfer is projected to state and local procurements and to the commercial sector.

Functional Area Support for R&D Phases

It has been pointed out that figure 1 represents opportunity for integration and teaming due to the multi-use of technologies and markets. Figure 1 also represents the opportunity for improvement through more efficient and effective use of functional support needed during the phases of R&D.

The product life-cycle (PLC), figure 3 is used as an educational and management tools to support overall planning functions associated with technology development. For success, the integration of functional support needs to be considered up-front and tailored throughout the product life-cycle.

The five traditional phases of the product life-cycle start with formalizing a need. This could be in the form of a new or improved operational capability, or a new or improved product, process, system or facility that has market opportunity.

The military has unique operational needs to meet mission requirements. For example: an operational need in the airforce might be for a new fighter airplane, F-xx, a new facility for treating hazardous wastes, improved systems and capabilities for improving the environmental quality of contaminated soils, improved capability for an existing depot/field level maintenance or specialized repair activity in areas; such as an electroplating industrial process, depainting, and recycling/reuse of waste oil.

Then how do the laboratories and other R&D organizations fit into the PLC for environmental needs? Figure 4 sets the stage for discussing the laboratory needs formalization and support provided in technology development. The technical base is continually being improved over time. In the environmental area these improvements and developments range from basic scientific and engineering technologies for contaminant removal, to operational, industrial and maintenance process wastes, to supporting developments of integrated and optimal systems. A critical part of the PLC is how development efforts evolve and transition.

At some point in the schedule for bringing in a new product or capability, the development organization integrates technologies from the technology base. This is not to imply that integration only occurs by acquisition organizations. For success, integration must occur at each phase of the PLC and by each organization involved in its development. To relate laboratory efforts with the PLC in figure 3, it is important to emphasize that, *"the degree to which laboratory efforts look downstream in the PLC when planning and performing basic, exploratory and advanced development needs, the more likely their products are likely to be used by acquiring organizations with the least risks and impacts to cost, schedule, and performance."*

Taking this a step further, lets look at the PLC from the view of an organization responsible for R&D and manufacture/construction to provide a market need or operational capability. Depending on risks assessment of the end product, this may need to be further broken down into phases of concept, validation and full-scale development or for facilities, this relates in-part to the percent complete of architectural design. It is through these stages that integration is done to trade-off and optimize all functional areas (operations, support and disposal) to best meet cost, schedule and performance objectives.

PLANNING

DEVELOPMENT
R&D PHASES

PRODUCTION
CONSTRUCTION

OPERATIONS
&
SUPPORT

NEED

DISPOSAL

Cost, Schedule, Performance Impacts

Figure 3. Product Life-Cycle

Products to Market

tech transfer/transition

Technology Base Development

program/acquisition offices

laboratories

industry

Figure 4. Technology Development.
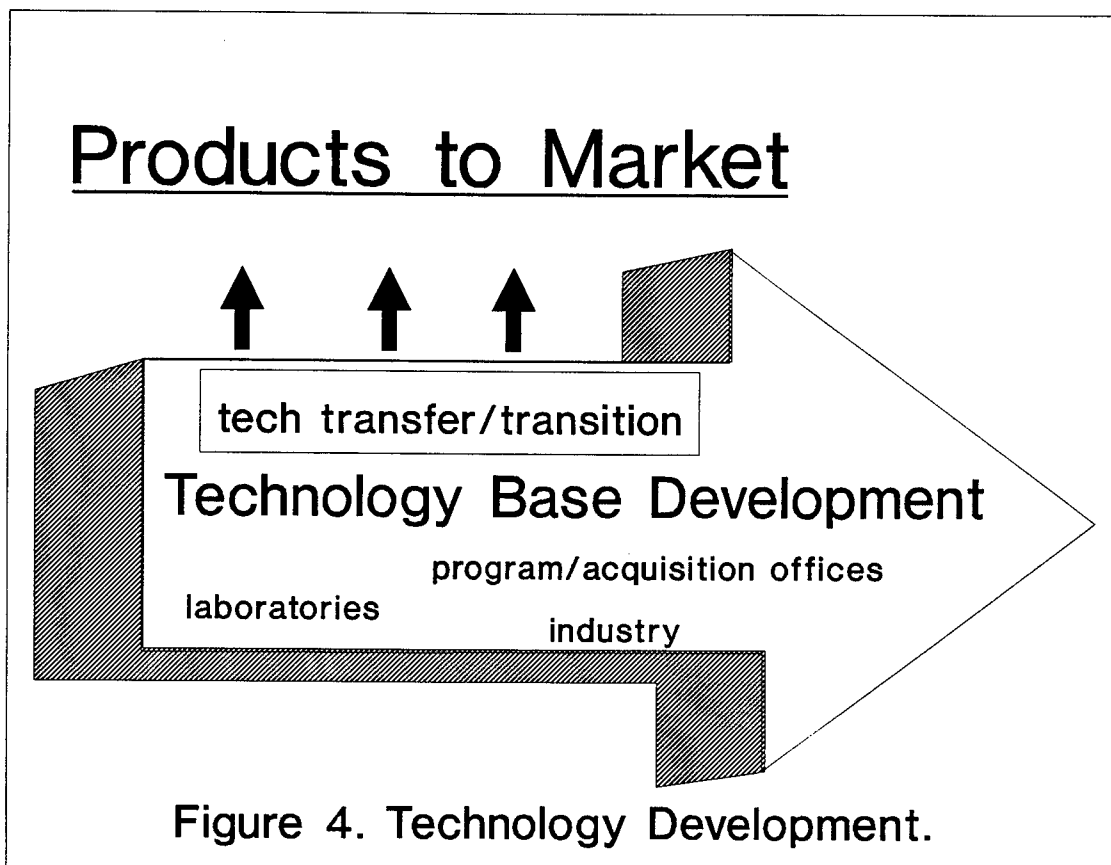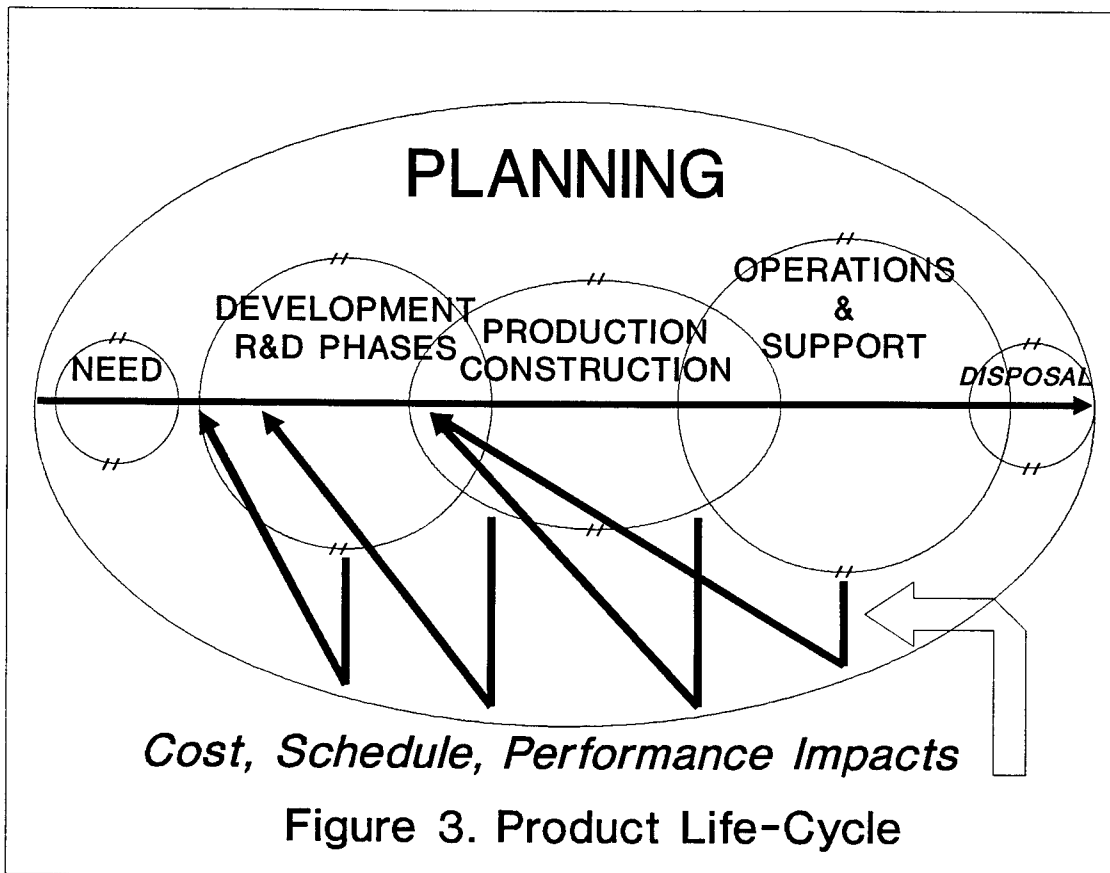
Thus during the development phase, particularly in the early stages of development, is where laboratory research and development efforts *"must be effective."* This is where the technology from laboratory efforts can either be a planned integration effort for transition directly into the acquisition program *(technology transition)*, or the technology has been used in the development and integration process by others as *(technology transfer)*.

In looking at the following examples, one can perhaps sense where technology development needs might be considered. In the example of a new F-xx landing gear or airframe system design, are their opportunities for material design trade-offs in performance, operations and maintenance, that include the environmental costs associated with the contractors industrial manufacturing processes and in support of operational units? Will the use of one industrial process or waste treatment for either manufacture and/or operations and maintenance support be more costly, considering the functional area of environmental considerations? Has the cost for disposing of the end product been adequately considered?

Is there a trade-off here for some essential performance parameters such as weight and speed of the weapon system? Will the use of one material for example: be lighter, but not last as long or be more difficult to repair, and present more environmental and total costs to the industrial processes for manufacture, downstream operational support and for its final disposition? At the other side of the spectrum; can a lighter material be manufactured that performs the intended mission, is easy to maintain and presents less environmental and total costs to the industrial processes for manufacture, downstream operational support and final disposition?

When one is to procure a new facility or technology to treatment or dispose of hazardous waste, it might be good to have a starting point showing flexibility for integration of the latest design and support concepts. This could include performance factors, such as; sizing, handling different wastestreams, other operational, support and political/community considerations, along with costs, benefits, and risks.

Are there benefits to benchmarking/baselining system capabilities for cost effectively improving the contaminants in soils? Is the projected market for soil improvement systems large enough in size and complexity where it would be cost effective to have functional and detail specifications for use in new requirements and to use as a baseline for providing improved capabilities?

Consideration for improving a maintenance capability, should come from benchmarking/baselining the current processes for its capabilities, costs, benefits and risks. What is the market for new technologies and in what timeframe? What should be the technology development thrusts be for electroplating or depainting processes? Does a particular recycle/reuse technology save costs to the organization? What will make it cost competitive for the organization? What degree of standardization and flexibilities for change are needed in providing operational capabilities?

It can be assumed that organizations with authority and resources (where-with-all) for acquiring new or improved operational capabilities (fielded systems/facilities), will perform their mission. The laboratories are one source for technology developments that other organizations rely on for obtaining environmental technology. Thus, part of the technology thrusts areas for

laboratories must be concentrated in "long-term projections," to be ready to support future acquisitions, market needs and improvement opportunities.

What types of basic and exploratory research should be supported to bring about the best bang-for-the-buck or return-on-investment (R.O.I.) in the long-run? Some could be high-risk high-pay-off technology thrust areas, and some might be a combination of moderate/low risk - high/moderate payoffs. What types of technology developments can allow the laboratory to be postured to support technology transition when planning of future acquisitions?

The third type of need is the case where the laboratory has been targeted as an integral part of the PLC for technology development including integration and "technology transition" to the acquiring organization. Thus, the laboratory involvement could range from a strong tie with the acquisition organizations in development of technologies that are scheduled for transition into the PLC, to technical transfer of industrial process and treatment technology information that could be considered during the integration process.

Another PLC integration area for improvement is in the planning of cost, schedule and performance considerations. It is generally recognized that all functional areas should be considered throughout the product life-cycle. Without this emphasis, the result is usually significant cost, schedule and performance impacts. Experience on most major efforts, even with emphasis in this area, shows a large opportunity for improvement. The arrows in figure 1 show what happens when the program is downstream and it is determined that a change is required. Just as the arrows are longer when one gets farther in the PLC, generally speaking the impacts to cost, schedule and potentially performance are increased. It could be very impactive to get to manufacturing or construction to determine that part of the operational

requirement was not included in the technology developed or that the operational need had changed. It could be very impactive to get to a manufacturing/construction decision and determine that consideration was not given to reliability, maintainability or logistics support. This also adds tremendous costs and schedule implications when long-lead procurement have been authorized and expended.

Specifically, when one forgets to plan up-front to consider potential environmental issues to be faced in the manufacturing or construction process, or the operations and support phase or when the products and supporting facilities are being phased out, then this could have tremendous impacts to initial cost and schedule estimates. It could be impactive to get a technology development approved based on a projection of "x" dollars, and to find out that the costs will now be "2x" dollars. It could be impactive to sell development based on a life-cycle cost savings of "x" dollars, but to find out that the savings will probably only be "x/2" dollars.

All functional areas and concerns throughout the PLC should be considered optimally in each phase of the PLC. For the environmental functional area this means planning for cost effective/competitive operational, industrial and maintenance process support and considering disposal of the product, system or capabilities. The challenge is in determining *"how"* to integrate the functional areas as-well-as *"how much"* for optimum consideration.

"RETURN-ON-INVESTMENT (R.O.I.)" - STRATEGY FOR RESOURCE PRIORITIZATION

Adopting R.O.I. as a strategy points to technology development opportunities. What are the quantitative and qualitative benefits for one technology development thrust versus another? What degree of emphasis is appropriate in up-front considerations of R.O.I. for one technology versus

another? This recommendation if to integrate comprehensive market analysis, cost, schedule and performance projections, risks and benefits analysis into a R.O.I. methodology for technology developments.

One aspect influencing the use of R.O.I. as a prioritization scheme is market analysis. Data on market potential and market share of a particular technology should include the particular D.o.D. component, for the D.o.D. overall, for other federal, state and local communities and for the international industrial base. The algorithm could allow different weighting factors ranging from total weight be for one organization or market category to equal weighting. An emerging characteristics coming from this strategy is to promote teaming and sharing of research and development costs based on the return potential. This allows a natural approach for communication and coordination among particularly D.o.D. and other Federal Agencies like the Department of Energy and the Environmental Protection Agency.

Another aspect with the use of R.O.I. is risk analysis. Risk analysis should consider both the risks associated with basic and applied technology base development. Risk analysis should include cost, schedule and performance risk projected to meet the projected return. Risks for some initiatives could be judged qualitatively into categories; such as, moderate to high and to others moderate-to-low in technical performance, schedule and cost risks. Concerning costs projections, range estimates identified to areas of uncertainty will allow an improved baseline for future technology management efforts during the product life-cycle. Each technology thrust will potentially have both qualitative, as-well-as, quantitative benefits. These may be political in nature, other functional areas such as safety, or might relate to potential markets. The point is that when considering the projected investment costs versus the amount of return sets up a

point of departure for prioritization of where to put resources. Adoption of this R.O.I. indicator can allow more objectivity in the prioritization process. Additionally, this baseline information can be valuable for use in program control, accountability, and technology transfer and transition purposes.

ADOPT "TOTAL PROCESS CONTROL" AS THE LONG-TERM TECHNOLOGY DEVELOPMENT OBJECTIVE

This recommendation is to consider adopting an objective for emphasis by R&D personnel that focusses on technology development. From the literature, it appears there are various goals and thrusts for the development and treatments of sludges. Depending on the perspective of the organization, these R&D thrusts range from goals of "total clean-up of superfund sites" and "Zero Discharge of Hazardous Waste by the Year 2000," to maximizing the development and commercialization of a specific technology or treatment process.

Although each of the above stated objectives may be appropriate for the particular level of emphasis, for the purpose of technology development, consider the top-level, long-range goal that fosters and emphasizes technology characteristics, *"to cost-competitively obtain the capability to restore the waste stream composition either to its original environmental characteristics or to another level that has been determined to be an improved state for the public's interest in either human health or the environment."*

For applications that have a waste stream, total process control can be considered the ideal technology development objective. This broad goal of total process control provides an important characteristics for emphasis that on the surface one might label as reuse or recycling.

However, this objective goes further by promoting the focus of technology development as one of process control. The technology(s) involved for total process control may evolve from some combination of the sciences and engineering disciplines. The point is to direct the technology development thrusts and emphasis toward that which will make significant long-term benefits to the environmental area.

ADOPT "COST COMPETITIVENESS" TECHNOLOGY DEVELOPMENT

The previous recommended objective includes the development of technology that will be "cost competitive." Challenging the technology development community to embrace "costs competitiveness" as a driver for technology development, can be considered a bold and needed change in technology development strategy.

In the environmental area, as in many areas, there is a tendency to mandate requirements and/or stated needs. The government may fund most of the research and development. Then, at the point, for implementation, cost effectiveness is the term used versus cost competitiveness.

What this cost competitive strategy does is to allow the government to change the paradigm for technology development to one that up-front in the product development cycle, stresses baselining and benchmarking of the current process needing change, and to develop the technology that will allow the process to provide the environmental need and still be competitive.

There is a subtle but critical difference between having a technology that helps to be competitive compared to seeking the most cost effective technology to meet a requirement. When one mandates a requirement or need, then the R&D community is charged during the product life-cycle to strive to insert cost effective analysis in trade-offs. The end result could

be the most cost effective alternative available, but the solution could present significant impacts to "government budgets" and not necessitate industry to "increase prices" on products out of the competitive range with international competitors.

SUMMARY/CONCLUSIONS

This study pointed to four areas that should provide significant improvement in the Air Force's environmental technology development program. Integration and teaming for needs formalizing and R&D planning will open up market opportunity and improve technology transition. Adopting R.O.I. as a prioritizing strategy will aid in the focussing on where to place resources. The long-range strategy of focussing on total process control as the long-term technology development objective can maximize R.O.I. in meeting future needs. Perhaps the most bold recommendation is to coordinate government activities and push the R&D community to target their efforts for environmental improvement without impacting the industrial competitive base. Although this study was for the Air Force, the recommendations apply to all federal, state, local community and industrial organizations engaged in environmental technology developments.

REFERENCES

[1]. Fact Sheet - Human Systems Center (AFMC) Armstrong Laboratory Environics Directorate, Office of Public Affairs.

[2] VISITT Vendor Information System for Innovative Treatment Technologies, User Manual (VISITT Version 2.0), EPA 542-R-93-001, Number 2, April 1993.

[3] ReOpt Remediation Software, Battelle Memorial Institute, D.o.E. contract, 1992.

[4] Managing New Technology Development, Wm.E. Souder and J. Daniel Sherman, McGraw Hill Engineering and Technology Management Series, 1994.

COMMUNICATIVE CHALLENGES FACING INTEGRATED PRODUCT TEAMS IN A HIERARCHICAL ORGANIZATION: A CASE STUDY

Malcolm R. Parks
Associate Professor
Dept. of Speech Communication, DL-15

University of Washington
Seattle, WA 98195

COMMUNICATIVE CHALLENGES FACING INTEGRATED PRODUCT TEAMS IN A HIERARCHICAL ORGANIZATION: A CASE STUDY

Malcolm R. Parks
Associate Professor
Dept. of Speech Communication, DL-15
University of Washington

Abstract

The special organizational challenges that face integrated product teams in a strongly hierarchical setting were explored. Military organizations such as the USAF exhibit high levels of vertical differentiation. That is, individuals and organizational units are related to one another hierarchically. The shift toward integrated product teams introduces a strongly contrasting model in which hierarchical relations are flattened and divisions based on function and technical specialty are crossed. This study represents an initial examination of the problems that arise when a unit based on integrated product teams (IPT's) is introduced into a strongly hierarchical organization. Specifically, this study focused on the Office for Prevention and Health Services Assessment (OPHSA) within the USAF. Three classes of organizational challenges were identified as part of a qualitative, participant-observation research project. These were: 1) intraorganizational conflict that arose when a new, flatter organizational structure was "grafted" into a larger, more hierarchical parent structure; 2) problems of cross-technical communication that arose when IPT members with different functional or technical specialties must work together; and 3) problems of role conflict and status ambiguity that arose within the new unit when members' tenure and rank were de-emphasized.

COMMUNICATIVE CHALLENGES FACING INTEGRATED PRODUCT TEAMS IN A HIERARCHICAL ORGANIZATION: A CASE STUDY

Malcolm R. Parks

Introduction

Advocates of "total quality management" (e.g., Bowen & Lawler, 1992) distinguish TQM from more traditional organizational practices in terms of its focus on the group rather than on the individual, its concern for egalitarianism, its belief in change rather than stability, and in terms of the value it places on participation, rather than command and control. The epitome of TQM thinking is the integrated product team (IPT) in which people are organized in flat, multidisciplinary teams around some specific function or problem.

Total quality management principles have penetrated large portions of the private sector and are now making their way into the Air Force. The USAF remains, of course, a strongly hierarchical organization. Nonetheless, examples of less hierarchically structured units, explicitly organized around the principles of TQM, are increasingly found as USAF not only voices the spirit of TQM, but also searches for creative ways to address new challenges. This can be clearly seen in the USAF medical mission. Changes in the world situation have brought new dimensions to the medical readiness mission such as a variety of new contingencies in military operations other than war. Changes in the shape of health care at home have spawned new challenges as the USAF moves toward managed care in a tri-service environment. Not surprisingly, then, new units have arisen to meet new needs. One of these is the Office for Prevention and Health Services Assessment (OPHSA).

By examining OPHSA we are afforded an opportunity to consider the special conflicts and challenges that arise when a relatively flat, egalitarian unit is grafted into a larger hierarchical organization. My research goal was to identify what these conflicts and challenges were. In order to set the context for my research my report opens with a brief discussion of the history and mission

of OPHSA. After a short section on methodology, I will describe the primary findings. The report

closes with a discussion of future research needs.

OPHSA's History and Mission

History. The Office for Prevention and Health Services Assessment (OPHSA) grew out of a widespread concern that the USAF lacked adequate outpatient and health surveillance data. These concerns date back at least to the early 1980's, but it was not until 1990 that it became apparent to the staff of the Air Force Surgeon General's office that new data would be required to address the issues that were looming over military health. These included changes in the readiness mission, the shift toward managed care, and initiatives such as Healthy People 2000 and Putting Prevention into Practice (PPIP). The seed that ultimately became OPHSA was planted in April, 1991 when the Surgeon General directed medical personnel in the San Antonio area to establish a working group to explore the feasibility of a Preventive Services Center. Although the model proposed by this group was rejected, the working group started a larger process that lead to the formation of a new study group in Armstrong Laboratory at Brooks AFB.

This group commissioned a study by Battelle to examine the needs and potential structure for a Preventive Services Center. Battelle recommended the formation of an IAC, an information analysis center, organized along the lines of already existing analysis centers. While an IAC was certainly responsive to some of the needs identified by the Surgeon General and Human Systems Center, it was not responsive to all of them. By mid-1993, it was agreed that a new unit with dedicated budget and personnel would be needed. A new working group was formed and directed to develop a Program Action Directive creating the new organization. This was to be an expanded organization that would include the assessment issues raised by the shift toward managed care. In November of 1993 the Air Force Surgeon General approved a budget and manpower documents. Final approval for the Program Action Directive was given in July, 1994 and the new office, now

renamed the Office for Prevention and Health Services Assessment or OPHSA, was set to start officially 1 October, 1994. The plan called for OPHSA to be a Directorate within Armstrong Laboratory at Brooks AFB in San Antonio. It was to be staffed by 12 officers, 8 enlisted personnel, and 17 civilians tentatively organized into three divisions. These were the Information Analysis and Studies Division, the Clinical Preventive Services Division, and the Executive Services Division.

Although OPHSA did not officially come into existence until the beginning of October, an implementation budget was in place prior to that and pieces of the new organization were beginning to come together by early 1994. Manpower destined for OPHSA was assigned to the Epidemiology Services Branch (AOES) at Armstrong Laboratory. This branch resides in the Epidemologic Research Division (AOE) of the Aerospace Medicine Directorate (AO) of Armstrong. By the Spring of 1994 a partially formed version of OPHSA was up and running within and alongside AOES. By August, a number of products were beginning to emerge under the banner of OPHSA. These included, for example, meetings with advisory boards that had been created to examine medical readiness and managed care. Even before they were officially ordained as a unit, members of OPHSA were preparing for their first meeting with the Executive Policy Board of the Air Force Surgeon General, the body to which OPHSA would be reporting.

Mission and Structure. The mission of OPHSA is to enhance the "capability to make evidence-based decisions" in Air Force medicine. It serves operational commanders and health program managers at all levels with improved measures and data to aid in decision making. More specific objectives include determining preventive and health services needs, developing clinical guidelines, improving the delivery of health services, creating and evaluating health promotion and other educational programs, conducting research that identifies areas for cost saving, and generally monitoring prevention and health research.[1]

To accomplish this mission OPHSA produces a limited range of products. That is, the actual work of OPHSA takes the form of consultations, literature reviews, technical reports, special

studies, publications, and technical conferences. While some projects can be done by a single individual (e.g., a simple consultation), most of OPHSA's projects will be done by an integrated product team. The vision statements and other statements to staff by Col Wright, the Director, and Lt Col Meyer, the Executive Manager, have consistently emphasized that OPHSA was to be organized around teams, that military rank would be secondary to substantive expertise, that roles would shift, and that quality management principles would be employed.[2]

Research Methodology

This research project involved a mix of qualitative, archival, and participant-observer methods. By using several different methods I was able to obtain a richer picture not only of OPHSA, but also of the environment in which it was situated in the Summer of 1994. Interviews with OPHSA members constituted the chief qualitative method. I was also given access to a wide range of documents pertaining to the history, functioning, and financing of OPHSA. Finally, I was an observer and a participant in the daily tasks of OPHSA. As such I not only observed, but also participated in the decision making. What follows is a summary of my primary findings.

Findings: Three Communicative Challenges Facing OPHSA

Based on the various methods noted above, I identified three challenges facing OPHSA as it moved through the Summer of 1994 and looked forward into its official start-up in October. These were: 1) intraorganizational conflicts; 2) problems of cross-technical communication; and 3) problems of role conflict and status ambiguity. Each of these challenges is described below along with a sampling of the strategies that were used to address them.

<u>Intraorganizational conflicts.</u> OPHSA is a new unit grafted into the existing structure of Armstrong Laboratory. By its very nature, a new organizational unit is an innovation that creates real and perceived threats to existing resource allocations and power balances within its host organization. These threats trigger responses that may influence the ability of the new unit to fulfill its mission and the ability of those in the new unit to fulfill their career objectives (See Brimm, 1992). My interviews and observations revealed numerous conflicts between OPHSA and its host organization. Owing to the sensitive nature of the conflicts, I will present only a brief delineation of them here.

The financial chain through which OPHSA receives funding has been a source of confusion and apparent envy to many in its host organization. OPHSA's funding came directly from the Surgeon General, but was passed down through the chain of command from the USAF Materiel Command to the Human Systems Division and to Armstrong Laboratory where it was passed through to OPHSA. One source of tension was the dictate that OPHSA funds where to be passed through the chain of command untouched. Thus, between the Surgeon General and OPHSA were a series of organizational units with accounting and pass-through responsibilities, but with no direct ability to influence the ultimate use of the funds. One of my informants compared this to the situation that would arise if one found that money over which one had no control was flowing in and out of his or her checking account. While this image no doubt oversimplifies the fiscal structures that fed OPHSA during the Summer of 1994, it squarely identifies the feelings of frustration and envy that were often seen among those in the chain of command above OPHSA. To make matters more confused, OPHSA was still residing within the Aerospace Medicine Directorate and was still functionally joined with Epidemiological Services in the Summer of 1994. Thus, not only was control over OPHSA lost, but it was also unclear where the boundary was drawn between OPHSA and other units for which the Aerospace Medicine Directorate (AO) did have direct responsibility. The result of this was a covert conflict in which considerable productivity was lost on all parts.

While it is tempting to suggest that this conflict should have been handled more explicitly (for example, by seeking interventions from the Surgeon General's office more often), my observation is that by keeping the conflict at a covert level the participants were able to save face and thereby avoid even worse conflicts. Ironically, the intraorganizational conflicts surrounding the establishment of OPHSA probably served to increase staff cohesion by creating a dim, but real, external threat.

A second source of tension revolved around housing and space. During the Summer of 1994, OPHSA was physically housed in the same building as other units. As new people came on board over the weeks space concerns became paramount, producing tension not only between OPHSA and its neighbors, but also between members of OPHSA who were forced to inhabit cramped, temporary quarters. It soon became apparent that the construction of a building to house OPHSA would be delayed considerably. The only available solution was to house OPHSA in two different buildings in the interim. This heightened tension still further because it raised issues of how OPHSA's staff would be divided. Strategies for coping with these tensions were quite limited. The director and executive manager sought to manage these uncertainties by making staff concerns a topic for open discussion and by prioritizing the establishment of computer links between the various offices that OPHSA would inhabit. Based on my research, I recommend that staff concerns about space and about being divided be aired regularly and nonjudgmentally at staff meetings. While these concerns will remain, their detrimental impact can be diffused by regular attention and acceptance.

A third source of intraorganizational conflict emerged once OPHSA moved toward the product development stage. Observations and interviews with those participating in OPHSA's advisory board meetings in August revealed that many people outside the organization view it as a mechanism for extending their own resources. They viewed OPHSA as a source of work that could not be funded through their own moneys. To some extent, OPHSA is indeed such a vehicle. It is

intended to serve the research and information needs of the USAF medical community. So these external expectations are integral to the primary business of OPHSA. On the other hand, it became apparent that many attending the advisory boards had unrealistic expectations for what OPHSA could accomplish. The challenge for OPHSA, like that of any new unit, is to establish control over its boundaries and resources in such a way as to build a broad constituency without also becoming purely reactive to an unending series of external expectations and demands. Two features of OPHSA's operation address this problem. One is the fact that its priorities are set on a quarterly basis by the Surgeon General's Executive Policy Board. Thus, requests for service from clients are required to pass through a definite external filter. The other feature is an internal filter. The Director and the Executive Manager have spent considerable time developing the role of "gatekeeper"-- that is, developing an internal protocol for referring requests for service. Thus, all requests for service are channeled through the Executive Manager who acts as a gatekeeper. This system was just being put in place during the time I was with OPHSA and it has yet to be fully tested. My interviews with staff suggest that making it work will be a trial and error process, especially once the staff is divided into different buildings where contact with the Executive Manager may be less immediate.

Problems of cross-technical communication. Cross-technical communication occurs when people from different functional groups must coordinate their activities and when people with different technical specialties must work together to produce a joint outcome. Efforts to communicate across these functional and technical boundaries may encounter the same problems of misunderstanding and hostility as efforts to communicate across cultural boundaries. In fact different technical areas may develop their own language, procedures, unwritten rules, and expectations.

Several features of OPHSA suggest that problems of cross-technical communication will arise. First, the overall design of OPHSA brings together distinct functional units within a

comparatively compact organizational structure. These units (e.g., executive services, information analysis and studies, clinical preventive services) will not only need to work together, but will need to work together in a constantly shifting ways. Second, these functional units will be staffed with people representing quite different technical specialties. Third, the work of OPHSA in intended to be innovative, thus requiring people to work together in entirely new directions. Finally, as a centralized resource for the entire USAF community, OPHSA will be required to address diverse client audience and needs. Concern for how technical information is "translated" from one audience to another is paramount under these conditions.

One of my research goals was to identify "critical linkages" in the cross-technical communication process. By this, I mean the places in the flow of information where problems of cross-technical communication were most likely to arise. Based on my interviews and observations, I believe that OPHSA will experience cross-technical communication problems at three particular points. The first of these is in the linkage between OPHSA staff researchers, USAF financial personnel, and contractors. Because much of the work of OPHSA will be done with contractors, it is essential that staff members be well informed about the legal and procedural elements of work with contractors. Current staffing plans call for hiring a contract administrator to act as a focal point in this process. The relationship between staff members and the contract administrator must be characterized by open and regular contact if problems of cross-technical communication are to be minimized. If the process is to work as envisioned, the contract administrator must be treated as a member of the integrated product team rather than as an isolated "answer man." Put another way, success at this critical linkage will only be obtained if the communication model is one of an on-going conversation rather than a question-and-answer session.

The second "critical linkage" in the cross-technical communication process within OPHSA will occur between the staff members engaged in specific research projects and the staff members whose primary technical specialties rest with the computerized databases and computer hardware.

To date, the development of OPHSA's database and computer capabilities has remained rather separate from the other activities of OPHSA. At the offsite in early September, for example, many of the staff members were genuinely uninformed about what was happening with the computer system. This problem is likely to worsen in the short term because of the divisions in housing. Many staff members will simply not be exposed to the computer people on a daily basis. Again, if problems of cross-technical communication are to be minimized, members on all sides must strive to develop regular, on-going contact. Researchers will benefit by taking time to educate computer specialists on the details of their projects and computer specialists will benefit by taking time to educate researchers about their resources and capabilities. In short, a priority should be placed both on formal training and on informal briefings. My judgment is that time that is initially lost on this process will be more than gained back by later increases in productivity.

The final "critical linkage" I wish to address was only indirectly suggested by my research. That is the linkage between OPHSA and it's USAF "public." In order to fulfill it's mission to be a resource for the broader USAF community, OPHSA must be able to communicate about its capabilities and products with that broader community. Currently, it is unclear what rules and conditions govern the distribution of information coming out of OPHSA.. I believe that the rules regarding distribution must be rather liberal if OPHSA is to fulfill it's mission. Otherwise, the timeliness and impact of OPHSA's work will be compromised. If problems are to be avoided, it is likely that policies covering information dissemination will need to be negotiated directly with the Executive Policy Board.

Problems of status and role ambiguity. If we cut through the rhetoric of integrated product teams to the practical level, a number of problems of status and role ambiguity become immediately apparent. Chief among these is the problem of military rank in a flattened team-oriented organization. OPHSA is envisioned to be made up of project teams in which rank is secondary to technical expertise. In a practical sense, this means that on some projects a Major might be

working on a team headed by a Captain. A Colonel might be reporting to a Lt Colonel. I observed both of these cases during the course of my research. The problems that accompanied these inversions of rank included openly expressed conflict, a hesitancy to give direction to a more highly ranked member of the team, and a tendency of staff members of greater rank to presume leadership roles that had not been given to them. But matters of rank are only one of the fault lines running through the efforts of the larger group to act cohesively. Another challenge to cohesion that I observed was the role shifts that occurred as personnel moved from project to project. A person who was a project leader in one case might simultaneously be cast in the role of an ordinary team member on another project. In addition, cohesion is challenged by the fact that project teams will include not only a mix of military personnel, but also a mix of civil service and contractor personnel. I observed cases, for example, in which contractor personnel could not be included in planning discussions because of potential conflicts of interest. Thus OPHSA is highly entropic with regard to issues of group cohesion. That is, cohesion is constantly dissipated by the changing nature of the tasks being done and by the partial suspension of role and rank as sources of role clarity.

The importance of addressing issues of cohesion is underscored by the relationship between innovativeness and group cohesion. The currency of OPHSA is new ideas. To be innovative, groups require a certain level of cohesion. Previous research has shown that innovativeness and group cohesion are positively related. In a study of 172 scientists, engineers, and technicians, for example, Scott and Bruce (1994) found that the individual's level of innovativeness was related to how positive workgroup relations were and how supportive managers were. Studies across a variety of organizations demonstrate that talk about innovations is more likely to occur between people who have a more personalized, multiplex relationship with one another than between people with weaker, impersonal relationships (e.g., Albrecht & Hall, 1991a, 1991b). In short, issues of status and role ambiguity must be addressed because they directly affect group cohesion which, in turn, has a direct influence on the level of creativity and innovation that OPHSA can bring to its mission.

Building and maintaining group cohesion requires on-going attention to the practical factors that build cohesiveness. These factors are generally known and shared by many different types of groups. One of the most practical lists of these factors comes from research on families (Settles, 1993). Seven interrelated "building blocks" of group stability and functioning have been identified: 1) shared space; 2) shared task and social activities; 3) common memories; 4) shared artifacts; 5) a strong group identity; 6) social support; and 7) rituals. Part of my role as a participant-observer was to share this information with OPHSA's members during the September offsite meetings. In addition, the Executive Manager has assigned a person to monitor the organization's cohesion on an on-going basis and to design activities that enhance the group's cohesion.

Future Research

My research was conducted early in the formation of OPHSA. It was during the period of time when the organization was moving from an implementation budget toward its regular budget. The organization occupied temporary facilities, was still partially wed to Epidemiological Services, and was not yet fully staffed. One avenue for future research, then, is to follow OPHSA's trajectory as an organization over time. Theoretically useful and practically valuable research could be conducted by examining how issues of intraorganizational conflict, cross-technical communication, and cohesion unfolded over time.

Another potentially fruitful line of research might examine the lifecycle of a single team project in some detail. Again, issues of conflict, cross-technical communication, and cohesion would be tracked. In this study, however, the goal would be to gain far more detailed picture of how these concerns were manifested in the workings of a single project team.

Finally, it would be useful to compare the workings of OPHSA with other organizations within the USAF that have adopted a flatter organizational structure and principles of TQM.

Through a comparative analysis, future planners would be in a better position to design and evaluate units that attempted to adopt a team-oriented structure within the broader hierarchical structure of the military.

Notes:

[1] Information compiled from OPHSA briefing documents developed by Col James Wright and Lt Col John Meyer.

[2] Information drawn from oral presentations made at OPHSA offsite meetings, September, 1994.

References:

Albrecht, T.L., & Hall, B. (1991a). Facilitating talk about new ideas: The role of personal relationships in organizational innovation. Communication Monographs, 58, 273-288.

Albrecht, T.L., & Hall, B. (1991b). Relational and content differences between elites and outsiders in innovation networks. Human Communication Research, 17, 535-561.

Bowen, D.E., & Lawler, E.E. (1992). Total quality-oriented human resources management. Organizational Dynamics, 20, 29-41.

Brimm, M.L. (1988). Risky business: Why sponsoring innovations may be hazardous to career health. Organizational Dynamics, 16, 28-41.

Scott, S.G., & Bruce, R.A. (1994). Determinants of innovative behavior: A path model of individual innovation in the workplace. Academy of Management Journal, 37, 580-607.

Settles, B.W. (1993). The illusion of stability in family life: The reality of change and mobility. Marriage and Family Review, 19, 5-29.

AURALLY DIRECTED SEARCH: A COMPARISON BETWEEN SYNTHESIZED AND NATURAL 3-D SOUND LOCALIZATION ENVIRONMENTS

David R. Perrott

Professor of Psychology and Director of the Psychoacoustics Laboratory

Department of Psychology

California State University Los Angeles

5151 State University Dr.

Los Angeles, CA 90032

AURALLY DIRECTED SEARCH: A COMPARISON BETWEEN SYNTHESIZED AND NATURAL 3-D SOUND LOCALIZATION ENVIRONMENTS

David R. Perrott
Professor of Psychology and Director of the Psychoacoustics Laboratory
Department of Psychology
California State University Los Angeles

Abstract

The present report describes the first two experiments of an extensive series that are currently being conducted on the application of spatial information derived from auditory signals upon visual processing: more specifically, this research investigates the impact of acoustic information upon a human subject's ability to locate and identify visual targets (maintenance of situational awareness). The results of the first experiment confirmed earlier reports (Perrott, Saberi, Brown and Strybel, 1990) that aurally directed visual search was substantially more efficient than unaided search even when the field to be scanned extended a full 360 degrees in azimuth and nearly a full 180 degrees in elevation. This baseline experiment was repeated with audio signals presented over earphones (a 3-D synthesized sound field). Performance in the latter situation was essentially identical to that encountered in the free field (i.e., natural environment), especially for visual targets initially located in the frontal hemi-field. These results indicate :(1) that free field listening environment can be generated in obviously non-free field situations (such as a cockpit of an aircraft) with little loss in the utility of the derived spatial input and (2) that such information can substantially improve the human subject's ability to process visual information.

AURALLY DIRECTED SEARCH: A COMPARISON BETWEEN SYNTHESIZED AND NATURAL 3-D SOUND LOCALIZATION ENVIRONMENTS

David R. Perrott

INTRODUCTION

A. Overview

In the design of aircraft cockpits, as is also true of most modern human work stations, it is fair to say that most of the information provided to the operator comes via the visual modality. And there are a number of excellent reasons in support of such a strategy. As Neisser (1967) pointed out, the visual channel can be thought of as a parallel processor without peer (at least with regards to other human sensory systems). Whether there is but one item present against a homogeneous background that forms the "field of view" or literally thousands of visual elements randomly arrayed across the same space, at one level the viewer can be said to have an awareness of all the items at the same time. But any attempt to extract more than the most rudimentary information from the array, say the shape of the red figures in the field, and it becomes immediately obvious that the visual system performs as a limited capacity processor restricted to but a small portion of the field of view. In effect, though an operator might see 20 illuminated dials displayed in front of him at the same time (parallel capacity), extraction of information from each dial can only be performed sequentially (serial capacity).

Technological advances over the last several decades has made it possible to greatly increase the quantity (and quality) of information that can be made available to the pilot. But while more information was expected to markedly increase the operators situational awareness and therefore performance, such expectations have not always been fully satisfied. The problem, of course, stems from the fact that processing of most visual information requires the pilot to attend to one small region of the viewing space until the data is extracted. Thus more information, no matter how useful, requires more processing time if each potential information source is to be utilized. It is probably no coincidence that the abrupt increase in information available to human operators parallels quite well the rapid and sustained growth in research dealing with what has been termed visual search.(see Wolf, 1994 for a review of this literature), a paradigm in which the subject is asked to find one item (a target) among a set of other items (distractors). Of course, in the context of a pilot operating an aircraft, the notion that one of many data sources should be examined (i.e., is a target) may not be evident without inspection.

By the 1980's, the emphasis had shifted from providing the pilot more information to helping him or her cope with the information already available. The idea that there in fact may be too much information available under some conditions via the visual channel (visual overload) prompted considerable attention on how to deal with this "new" problem. A number of

approaches were tried with varying degrees of success, a reconsideration of how the visual information is displayed, for example (e.g., head-up-displays and helmet mounted displays).

One approach to the problem of "information overload" that has been gaining increased attention begins with the counter-intuitive notion that the operator of a complex system might optimize the utilization of the existing array of information with the addition of yet more information. Numerous researchers recognized that the bottleneck suffered by the system operator was not too much information but rather too little information as to which information source was most relevant at any given moment. The potential of providing the missing information via the auditory channel was quickly identified (e.g., Doll, et. al., 1986).

By the mid-1980s, the Human Engineering Division of the Air Force Aerospace Medical Research Laboratory at Wright Patterson Air Force Base began a program of research to develop a system for the synthesis of a "free field" listening environment (McKinley, 1988, Ericson and McKinley, 1989 and McKinley, Ericson and D'Angelo, 1994). The plan was to provide auditory spatial information to human operators under conditions that were not normally conducive to such an attempt (e.g., the cockpit of an aircraft) by recreating the information normally available to a listener in a free field via earphones. It was felt that the successful simulation of a 3-D auditory array could improve operator performance by increasing the operator's "operational awareness". Exactly how much advantage could be obtained by providing auditory spatial information was unknown when this program was initiated.

During this same period, the Psychoacoustic Laboratory at California State University, Los Angeles also started a long term project directed at examining what use human subjects could make of spatial information derived from the auditory modality. For example, while there was an extensive literature that existed regarding the ability of humans to utilize spatial information from the visual modality to direct behavior (e.g., eye-hand coordination, Woodworth & Schlosberg, 1954), no systematic attempt had been made with spatial information derived from the auditory modality (i.e., ear-hand coordination). Indeed, what work had been reported tended to focus upon the inverse issue, that is whether movements by listeners would alter auditory localization performance.

It eventually became obvious to the participants in these laboratories that the two independent research programs were at least complimentary in nature. For example, some of the early findings obtained in the free field (e.g., Perrott, 1988a and 1988b) suggested that a substantial improvement in visual information processing could be achieved if spatial input from the auditory modality were made available but, outside of the laboratory, the listening conditions in most "modem" environments were seldom conducive to localizing sounds. It seemed clear that the 3-D sound system that the Air Force was developing could greatly expand the number of situations and tasks that could benefit from the application of this data. Similarly, while the Air

Force research effort (now the Bioacoustics Branch of the Armstrong Aerospace Medical Research Laboratory) had made considerable progress in the development of their 3-D auditory system a substantial program of research was still needed to determine just how effective their system was in the "simulation" of free field localization cues and, equally important, what advantage could be achieved if such information was made available to a human operator.

The experiments that are reported in this paper represent the initial attempt to combine the efforts of these two laboratories toward a common goal: An assessment of the impact of auditory spatial information upon the information processing capacity of a human subject. In the first experiment, the effects of the presence or absence of auditory spatial information upon a visual search task was examined in an anechoic environment (i.e., under ideal free field listening conditions). This would be the referent condition since almost all naturally occurring auditory spatial cues would be available to the subjects in this situation. The second experiment, using the same subjects, was identical to the first except that the auditory spatial information was delivered through earphones using the 3-D sound system developed by the Air Force. The rational for the particular experimental paradigm employed in both experiments is developed in the following sections.

B. Auditory Psychomotor Coordination

As noted above, the Psychoacoustics Laboratory at California State University Los Angeles began a long term program of research concerned with the utilization of the spatial information acquired via the auditory modality. In the initial series of experiments (see Perrott, Ambarsoom and Tucker, 1987), the focus was upon auditory psychomotor coordination or more specifically, the ability of subjects to regulate spatially organized behavior based upon spatial information from the auditory modality. The first case that was considered, having a subject turn his head so as to "face" a sound source, was stimulated by the fact that such a shift in orientation in response to the onset of a sound had been frequently observed in man and animals (e.g., Pavlov, 1927; and Sokolov, 1967). As by way of example, in a seminal series of publications, Konishi and his co-workers measured localization performance in the barn owl, an organism that utilizes sounds in its nightly search for prey. Observation of the head position, in this specie, proved to be both a reliable and sensitive measure of localization capacity.

However, the attempt to extend this approach to human subjects (Perrott, et. al., 1987) was not particularly successful. Human subjects pointed their nose at sound sources with considerably less accuracy than they could discriminate the position of a sound source using a non-motor response. It eventually became evident why college sophomores do not behave like owls. The barn owl must turn its face toward a sound source because the eyes are essentially immobile. College sophomores, on the other hand, are not so constrained. An analysis of video

tapes obtained while the subjects were localizing a hidden sound source revealed that concurrent movements of the head and eyes were typical when a subject was asked to "face" a sound source.

Eventually it became clear that, in spite of the instructions, "straight ahead" for these subjects was, within broad limits, defined by where their eyes were directed. While it was clear that one should be able to train a subject to point his nose at a sound source, the fact that shifts in gaze toward the source seemed far more natural and, upon reflection, a far more useful response for a human to make forced a reconsideration of our general approach. Prompted by these observations, a series of experiments was initiated to evaluate just how well subjects could point their eyes at sound sources.

C. Visual Search

While the visual modality has exceptional capacity to resolve the distribution of light, the fine resolution required to correctly identify say the letter "A" on this page extends, at most, only a few degrees from the line of gaze (the image must fall on or near the fovea). Reading or any other activity that requires a relatively high degree of acuity requires that the subject frequently make adjustments in the position of her eyes so that the energy to be evaluated falls in the central visual field. The eye (saccadic eye movements) and head movements that are encountered when human subjects attempt to fixate upon a new visual target can be quite fast. Indeed, shifts in the line of gaze in excess of 700 degrees per second may be encountered as the subject moves to a new fixation point. In effect, little time is required to change from one fixation point to another in the immediate field, though the time required to initiate a head and/or eye saccade is an entirely different matter (latencies on the order of several hundred ms are commonly encountered, see Perrott, et. al., 1990).

In a series of papers (Perrott, et. al., 1987; Perrott, 1988b; and Perrott and Saberi, 1988) the following speculation was offered: To understand modern human auditory localization performance, which appears to be relatively good compared to some species that have been studied, it might be useful to consider this capacity from an evolutionary or ecological perspective. In humans, the forward placement of the eyes is essential for an extensive binocular field of view (and excellent depth perception) however, the resulting binocular capacity in this relatively large headed animal came at some potential expense: At all times, more than half of the immediate environment is out of the field of view. Modern humans, along with other species that have to observe the world from a "narrow window", probably spend an inordinate amount of time just scanning the immediate environment. Of course, under natural conditions (unlike the modern environment), our ancestors had to be concerned with predators and potential prey as they moved around the environment. The absence of visual capacity in the rear hemi-

field is a significant limitation. It was in this context that we argued that the ability to resolve the location of a sound source has a critical role, at least for humans.

While most objects in the world are normally quiet, those that emit sounds are frequently significant. For a specie with a restricted view, a re-orientation of the gaze toward the source of sound would make sense. Indeed, for both humans and dogs, two species that have similar visual constraints, the orientation toward the source of a sound is so common as to have been labeled a "reflex" (see Sokolov, 1967 for a discussion of the orientation reflex). When considered in this context, the auditory spatial channel could be said to have a significant role in the determining of what information the visual spatial channel processes.

This capacity of auditory spatial channel to "control" the visual modality is particularly evident if the sound is intense or unexpected or even just novel. Such events seem to demand "attention" and can readily disrupt otherwise significant ongoing behavior. Thus there were numerous reasons to expect that spatial information from the auditory channel, when used to "point" the visual channel to relevant events should be successful.

The results from several experiments concerned with the impact of spatially correlated sounds that simply indicated "where" visual information could be obtained (Perrott, 1988a; Perrott, 1988b; Perrott, et. al., 1990; and Perrott, et. al, 1991) clearly supported this expectation. Two functions were identified and will be discussed here. The first involves the ability of a subject to discern that a change (any change) has taken place in the environment. Aside from the obvious advantage that auditory events can be detected regardless of the relative orientation of the subject at the moment the event occurs, it is also well known that with similar stimulus levels, simple reaction times to sounds are faster than those obtained to lights. A latency advantage on the order of 20-40 ms (e.g., Woodworth and Schlosberg, 1954), a relatively small effect, has been reported for auditory stimuli when compared to the latencies obtained for visual events located in the central visual field. However, when more peripheral locations (out to 80 degrees from the fovea) are considered, the auditory advantage can expand to a hundred ms or more even when the subjects are only required to report that an event has occurred (Sadralodabai, Cisneros and Perrott, 1994).

The second function is concerned with the accuracy with which shifts in "gaze" can be directed toward visual targets. As noted earlier, such shifts can be completed very quickly. But, due to the long latency required to organize such a movement, errors in the movement (i.e., the shift fails to bring the target on to the fovea) are particularly costly since an additional interval is required to organize the additional movement (an intersaccade latency). Under conditions in which a single visual event is presented within the visual field, the reduction in the time to localize and identify which of two targets is present can be as great as 500-700 ms if a sound source is present from the same location as the visual target (the sound contains no information

as to which target is present, only where it is located). Such large effects were evident for targets initially located within the visual field, between 60-80 degrees from the fovea (Perrott, et. al., 1990). It was the magnitude of this latter effect that led to the hypothesis that localization accuracy was substantially enhanced by the presence of spatial information from the auditory modality. More recent research has tended to confirmed this explanation. In the periphery, in excess of 25 degrees from the fovea, auditory localization performance is moderately superior to that observed in the visual modality (Perrott, et. al., 1993) and the best localization performance in the periphery is observed when both auditory and visual spatial information are available concurrently (Perrott, 1993).

All of the previous research that has been performed involving visual search in the presence of spatial information from the auditory modality has been confined largely to the frontal hemi-field (within 120 degrees of the initial line of gaze) and generally to only limited variations in the possible elevation of the visual target relative to the original fixation point maintained by the subject. In the first experiment, the time required to localize and identify which of two visual targets was presented on each trial was determined for events broadly distributed across both the front and rear hemi-fields.

I. EXPERIMENT 1. LOCALIZATION AND IDENTIFICATION OF VISUAL TARGETS: EFFECTS OF AUDITORY SPATIAL CUEING IN THE FREE FIELD

A. Methods

Five subjects, ages 20-25, participated in all aspects of the experimental program. Four of the five were drawn from the experimental subject pool maintained by the Armstrong Laboratory and the fifth was one of the authors (J.C.). None of the subjects reported any history of visual or auditory abnormality.

The tests were conducted with the subject's head located at the center of a spherical array of loudspeakers (radius of 7 feet). While a total of 272 loudspeakers would be required to provide sources spaced evenly at approximately 15 degree intervals across this entire space, the 10 locations directly below the subject were not used in these experiments. The speakers, in turn, were mounted on the inside of a geodesic sphere, the surface of which was covered with acoustic foam to reduce reflections. And finally, the whole apparatus in which the subject was seated was mounted inside a large anechoic chamber.

Mounted on the front of each of the 262 loudspeakers was a four element array of L.E.D.'s distributed to form a diamond. During testing, either the center two elements of the array from one speaker would be activated on a given trial to form a vertical "line" or the lateral two L.E.D.'s would be active forming a horizontal line. The relatively small size of the figures that were generated by this arrangement was done to ensure that the subjects would have to

employ the central visual field to discriminate which was present. Similarly, the use of two active L.E.D.'s in both configurations avoided any systematic difference in brightness that might allow the subjects to identify whether the vertical or horizontal configuration was present without having to bring the target into the central visual field. The L.E.D. displays produced a moderate amount of illumination (16 foot-candles) that was readily apparent in the otherwise dark chamber. Under the circumstances employed, the visual targets, when activated, provided a high contrast source that could be readily detected.

Under all conditions tested, the primary task was a two-alternative, force-choice paradigm in which the subject had to indicate, via push button, which of two visual targets was present on that trial (a vertical or horizontal "line"). Similarly, all tests were conducted with a high degree of spatial uncertainty as to "where" the next event would occur. Within a session, all 260 locations from which a visual target might be generated had an equal likelihood of being selected on a given trial. Since the visual target could be readily identified once the image had been brought within the central visual field (few identification errors were expected), the primary problem faced by the subject was "finding" the target and making the appropriate shift in the line of gaze.

Two experimental conditions were evaluated across successive blocks of trials (five sessions per subject per condition). In the Spatially Uncorrelated Sound Condition, all trials began with the onset of a broad band noise presented at from a speaker at a fixed location. The sound was used to indicate that a visual target was now active and that the subject should begin his search. The sound did not provide any information regarding "where" the target was located. In the second condition, the Spatially Correlated Sound Condition, the same general configuration was employed except that sound now came from the same location as the visual target (one of 262 speakers-target locations). All other aspects of the experimental sessions were identical.

The subject was instructed to remain seated and facing forward (zero degrees azimuth and zero degrees elevation was defined as that point directly in front of the subject's nose) during the three second inter-trial interval (i.e., the head was only approximately centered in the speaker array since we did not wish to restrict or alter the subject's natural movements of head and or body with a bite bar or other such device) until the auditory cue was sounded. At the onset of the acoustic signal, they were instructed to locate and identify as quickly as possible whether the two L.E.D. array was in the horizontal or vertical configuration. In actuality, they were encouraged to move their eyes, head and even torso in whatever manner seemed natural while they searched for the visual target. And finally, having located the target they were to indicate which array was present on that trial by pushing one of two hand held buttons. Their response was used to terminate both the visual target and the acoustic cue and begin a new

inter-trial interval. They were then required to return to their initial position (facing forward) to await the next trial. Targets were presented from all locations within each session (262 trials) using a randomization without replacement technique. Both the elapsed time between the onset of the visual target and the subject's response (reaction time) and whether or not his response was correct was recorded for each of the 262 target locations. Each subject completed all 5 replications of a given experimental condition before continuing on with the next task (fixed-block design), though the order that the subjects performed each experimental condition was randomized.

B. Results

The latencies obtained when the subjects were required to locate and identify which of two visual targets were present as a function of the location of the target relative to the subject's initial line of gaze, without benefit of spatial information from the auditory modality were as follows: For most of the frontal field, latencies fell within the range of 1000-1500 ms. A rapid increase in latencies was encountered beyond roughly 80 degrees azimuth and, to a lesser extent, above or below 50 degrees elevation (an increase of approximately 1500 ms). These results were not unexpected since the targets were now at or just beyond the limits of the subject's field of view. For visual targets located in the rear hemi-field, latencies in excess of 3000-5000 ms were common (3-4 times those observed for targets in the central visual field).

The advantage obtained when the locus of the visual target is marked by a sound source at the same location was clearly evident. For most targets located in the frontal hemi-field, latencies are generally less than 1000 ms. But probably the most remarkable effect is evident in the rear hemi-field. Latencies were consistently below 1500 ms for almost all locations examined. In effect, in the presence of spatial information from the auditory system, subjects can locate and identify visual targets behind them with the same efficiency as they could locate "uncued" targets in the front.

If one only considers the relative distance of the target from the initial fixation point in terms of azimuth, most of the advantage created by informing the subject with a spatially correlated sound regarding where to find the visual target is evident in the rear hemi-field. The reduction in search time is in excess of 2500 ms for most of the latter target locations.

In contrast, if one only considers the relative elevation of the target, the largest reductions in search time occur with events located 50 degrees below the initial line of gaze (a saving of 2000 ms or more). But a reduction on the order of 1000 ms, a substantial improvement, was clearly apparent across the remaining elevations. There seems little question that the subjects were able to utilize the auditory spatial information to resolve the relative elevation of the target.

II. EXPERIMENT 2. LOCALIZATION AND IDENTIFICATION OF VISUAL TARGETS: EFFECTS OF AUDITORY SPATIAL CUEING WITH A 3-D VIRTUAL SOUND SYSTEM

A. Methods

All aspects of this experiment are the same as the proceeding experiment except that the spatially correlated sound employed to direct the subject to the visual target was generated by the Air Force version of a 3-D virtual sound display yoked to a head-tracking device. An extensive technical discussion of this system can be found in McKinley (1988).

B. Results

A summary of the performance obtained when the locations of the visual targets were cued using the virtual sound system is as follows: In general, performance is very similar to that encountered when actual sound sources were used for targets located within 130 degrees of the initial line of gaze. For most of the frontal hemi-field reaction times are less than 1000 ms and for targets located just beyond the limits of the initial visual field (80-130 degrees) the range increases to 1500 ms. At greater azimuths, the virtual sound is both less effective than the non-simulated and more effective than the spatially uncorrelated sound condition.

The reduction in the latencies generated by the virtual sound system relative to the un-cued condition is remarkably similar to that describe earlier when sounds were presented in the free field. Greatest improvement is evident in the rear hemi field (azimuths greater than 90 degrees) and at the lower elevations (below 50 degrees).

Some "cost" was encountered by our subjects when the auditory spatial cue was delivered by the 3-D virtual sound display (re. natural free field listening conditions). In general, latencies are longer when the cues are presented over earphones however, the performance is exceptionally good (i.e., similar to that with "real" sources) within 90 degrees of the initial fixation point. In terms of azimuth, localizing targets in the rear hemi-field required an additional 500 ms to be accomplished. And in terms of variations in elevation, the latencies are some what longer (several hundred ms) when the auditory cues are delivered via earphones.

III. Discussion

The argument that I would like to make is that the spatial channel of the auditory system evolved, in humans at least, to serve the ocular motor system responsible for shifts in gaze. While the results of the first experiment are in complete agreement with this proposition, I recognize that this "evolutionary" hypothesis is not directly testable. What I can say is that the localization and identification of visual targets can be completed far more quickly when spatial information from the auditory modality is provided. The improvement obtained for events in the frontal hemi-field replicate our earlier observations (e.g., Perrott, et. al., 1990) and the impact of

this information for events in the rear hemi-field seems to provide a reasonable extension of this earlier research.

I also believe that the improvement in visual target acquisition obtained here represents a minimum description of the advantage available. First, as noted earlier, signaling the subject to begin the search for a target using a spatially uncorrelated auditory event (the control condition used here) is more effective than if a visual signal had been used. Indeed, without the spatially uncorrelated sound, latencies would have been considerably longer in the peripheral regions of the frontal hemi-field (Sadralodabai, Cisneros and Perrott, 1994) and, of course, the task would have been nearly unmanageable for the subjects for events out side of the initial field of view. And second, all tests were conducted under low illumination and without the presence of "visual distracters". Both of the latter aspects would make the task of detecting and localizing the visual target using only spatial information from the visual channel considerably easier. In the low ambient light available, the visual target array stood out as a singular, well defined figure and not merely one item from an array of potential visual targets. As has been demonstrated in earlier research (Perrott, et. al., 1991), the advantage created by providing spatially correlated sounds increases substantially as a direct function of the number of alternative visual figures (distracters) present in the field. In summary, I believe that the current test provides a conservative estimate of the value of auditory localization cues in directing gaze.

The results from the second experiment are particularly encouraging. Before I attempt to identify the limitations of the Air Force's system used to "simulate" free field auditory spatial information, let us start with a description of the system's successes. First, when compared with the control condition (spatially uncorrelated sound), the simulation of the auditory spatial information markedly improved visual search performance, sometimes by several seconds, regardless of the initial location of the visual target. And second, even when compared to spatially correlated sounds in the free field, for events in the forward hemi-field (extending approximately 80 degrees laterally in both directions and 40 degrees vertically) search latencies were essentially the same in these two conditions. Of course, this is the region of greatest concern. Whether one provides information using a heads-up-display (HUD) or the more common instrument panel and cockpit windows, most of the critical signals will tend to be located in the frontal hemi-field.

As noted earlier, the "simulation" was systematically less effective for targets located more than 110 degrees from the initial fixation point. Part of the explanation may lie in the rate at which the system can up-date information to the subject. With the large movements required to orient toward a source say at 180 degrees azimuth (head and torso), "average" velocities in excess of 300 degrees per second did occur and "peak" velocities well above even this rate would be common. Thus, unlike the condition encountered with real sources, during particularly

rapid movements, incorrect information regarding the "current" locus of the target would be given to the subject.

The performance obtained as a function of the relative elevation of the visual targets was also degraded relative to that obtained in the free field listening condition. The fact that individual head-related-transfer functions were not employed would seem to be a reasonable explanation for this "failure".

In conclusion, the application of spatial information from the auditory modality does generate significant advantage for human subjects attempting to locate and identify visual targets, for all regions of the subject's immediate field. Moreover, the simulation of a 3-D auditory display seems to be a practical method by which such information can be made available to human operators regardless of the characteristics of the environment in which they are located. And more specifically, the 3-D auditory space simulation system developed at the Armstrong Laboratory over the last decade can readily be used in that capacity with significant benefit.

REFERENCES

Doll, T., Gerth, J., Engelman, W. and Folds, D. (1986) Development of simulated directional audio for cockpit applications (USAF Report AAMRL-TR-86-014). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Laboratory.

Ericson, M. and McKinley, R. (1989) Auditory localization cue synthesis and human performance. Proceedings of the National Aerospace and Electronics Conference. Pgs. 718-725.

McKinley, R., Ericson, M. and D'Angelo, W. (1994) 3-dimensional auditory displays: Development, applications and performance. Aviation, Space, and Environmental Medicine, 65, A31-A38.

McKinley, R. (1988) Concept and design of an auditory localization cue synthesizer. [Thesis] Air Force Institute of Technology; AFIT/GE/ENG/88D-29.

Neisser, I. (1967) Cognitive Psychology. New York: Appleton, Century, Crofts.

Pavlov, I. (1927) Conditioned Reflexes (G.V. Anrep, Trans.). New York: Oxford University Press.

Perrott, D. (1988a) Auditory psychomotor coordination: Auditory spatial information can facilitate the localization of visual targets. Proceedings of the Localization Symposium, (CHABA), Washington D.C.

Perrott, D. (1988b) Auditory psychomotor coordination. Proceedings of the Human Factors Society 32th annual Meeting (pp. 81-85). Santa Monica, CA: Human Factors Society.

Perrott, D., Ambarsoom, H. and Tucker, J. (1987) Changes in head position as a measure of auditory localization performance: Auditory psychomotor coordination under monaural and binaural listening conditions. J. Acoustical Society of America, 82, 1637-1645.

Perrott, D. and Saberi, K. (1990) Functional integration of auditory and visual sensory processes. Paper presented at the Annual Meeting of the American Psychological Society, New Orleans.

Perrott, D., Saberi, K., Brown, K. and Strybel, T. (1990) Auditory psychomotor coordination and visual search behavior. Perception and Psychophysics, 48, 214-226.

Perrott, D., Sadralodabai, T., Saberi, K. and Strybel, T. (1991) Aurally aided search in the central visual field: Effects of visual load and visual enhancement of the target. Human Factors, 33, 389-400.

Perrott, D. (1993) Two modalities and one world. Proceedings of the International Meeting of the Audio Engineering Society. Copenhagen, Denmark.

Perrott, D., Costantino, B. and Cisneros, J. (1993) Auditory and visual localization performance in a sequential discrimination task. J. Acoustical Society of America, 93, 2134-2138.

Sadralodabai, T. and Perrott, D. (1990) Effects of spatially correlated auditory information on eye movements. Paper presented at the Fall Meeting of the Acoustical Society of America.

Sokolov, Y. (1967) Perception and the Conditioned Reflex (S.W. Wayenfeld, Trans.) New York: Pergamon.

Wolf, J. (1994) Guided search 2.0: A revised model of visual search. Psychonomic Bulletin and Review, 1, 202-238.

Woodworth, R. and Schlosberg, H. (1954) Experimental Psychology. New York: Holt, Rinehart and Winston.

INTERACTION BETWEEN NITRIC OXIDE AND OXYGEN
DURING HYPERBARIC OXYGENATION

Edward H. Piepmeier, Jr.
Assistant Professor
College of Pharmacy

University of South Carolina
Columbia, SC 29208

Final Report for:
Summer Research Program
Armstrong Laboratory

DOSE RESPONSE STUDIES FOR HYPERBARIC OXYGENATION THERAPY

Edward H. Piepmeier, Jr.
Assistant Professor
College of Pharmacy
University of South Carolina

Abstract

The primary goal of this summer's research was to identify appropriate dose-response studies as a basis for the determination of optimal hyperbaric oxygenation therapy regimens. Several specific responses to hyperbaric oxygenation therapy were examined including; white blood cell surface protein expression, nitrite production and apoptosis. Preliminary results indicate that hyperbaric oxygenation may be utilized to both suppress deleterious immune responses while enhancing advantageous immune response.

DOSE RESPONSE STUDIES FOR HYPERBARIC OXYGENATION THERAPY

Edward H. Piepmeier, Jr.

INTRODUCTION

Dose response studies are currently required by the Food and Drug Administration for all new drugs. These studies are conducted to identify the dose of a drug which is necessary to elicit the response (indication) for which the drug is being marketed. In addition, dose response studies identify side effects which may be beneficial or detrimental to the patient. Occasionally, a side effect that is found in a dose response study becomes the basis for another indication for the drug. The use of minoxidil for hair growth and the use of aspirin for the prevention of platelet aggregation are examples of indications which were originally side effects.

When identifying the optimal dose of a drug, it is critical to identify at what point the damage from the side effects outweighs the increased benefits of the indicated effects. A dose which does not elicit the desired response gives a blood level which is below the minimum effective concentration. A dose which elicits undesired responses gives a blood level which is above the minimum toxic concentration. Doses which give drug concentrations which fall between the minimum effective concentration and the minimum toxic concentration are defined as therapeutic. Because most drugs have both multiple desired and multiple undesired effects, these effects must be carefully considered in the treatment of each patient. For example, an undesired effect such as dizziness may be preferable to pain when identifying the optimal dose of an analgesic.

Hyperbaric oxygenation therapy has been approved for use in: air or gas embolism, carbon monoxide poisoning and smoke inhalation, carbon monoxide poisoning complicated by cyanide poisoning, clostridial myonecrosis, crush injury, compartment syndrome, and other acute traumatic ischemias, decompression sickness, enhancement of healing in selected problem wounds, exceptional blood loss anemia, necrotizing soft tissue infections, refractory

osteomyelitis, radiation soft tissue damage and osteoradionecrosis, compromised skin grafts and flaps, and thermal burns. The areas currently under investigation for hyperbaric oxygen treatment include; blunt force injuries, penetrating injuries, open fractures and bone grafting, chemical thermal burns, ionizing/nonionizing radiation injuries, skin grafts and flaps in compromised tissue, nerve regeneration, aseptic bone necrosis, organophosphate poisoning, septic shock, cyanide poisoning, cerebral edema, cold-frostbite, spinal cord injury and explosive decompression. However, like aspirin which was used for thousands of years without knowing its mechanism of action, the mechanism of action for hyperbaric oxygen therapy in many of these disease states is unclear. Only twenty years ago aspirin was identified as a cyclooxygenase inhibitor and it has only been within the past few years that some of the immune effects of hyperbaric oxygenation therapy have been identified. This study further defines the immune effects of hyperbaric oxygenation treatment which are responsible for the therapeutic use in the treatment of many of the above indicated disease states. In addition, results obtained in support of these preliminary studies could be used to identify the potential of hyperbaric oxygenation therapy in other disease states. The dose response studies which follow these preliminary studies conducted this summer will likely produce a plethora of additional indications.

METHODS

Phenotype Labelling - Phenotyping for CD4+(T-helper)/CD8+(T-suppressor) cells was effected with the Becton-Dickinson Simultest fluorescent antibody mixture. Anti-CD4+ was fluoroisothiocyanate conjugated; anti-CD8+ was phycoerythrin conjugated. The cells were incubated on ice for thirty minutes with one microliter of the fluorescent antibody mixture. Statistics were obtained with quadrant analysis of dot plots on FACScan Analyzer. Phenotypic analyses of other surface markers were made in a similar way using the appropriate

fluorescent-conjugated primary antibody. Becton-Dickinson monoclonal antibodies used included: Leu 7-PE recognizing CD-57 for natural killer cells; Leu M3-PE for monocytes and macrophages CD-14; Leu 12-PE for B cells; Anti-IL2R FITC conjugates for CD25 surface antigen.

Flow Cytometry - Cell phenotype analyses were performed with FACScan Flow Cytometer (Becton-Dickinson, San Jose, CA.) Cells were excited with a single 488 nm argon laser and 10,000 events were collected. Data was analyzed with the Becton-Dickinson Lysis Program. Autofluorescence was estimated using dot plots with forward scatter versus side scatter.

Cell Preparations - Sterile cell culture techniques were followed. Peripheral blood mononuclear cells were isolated by Ficoll Density Gradient Centrifugation. Cells were cultured in RPMI medium supplemented with 1% penicillin-streptomycin and 10% heat inactivated fetal calf serum.

Wound Cells - Bandages were removed from diabetic patients with non-healing wounds. The bandages were washed with phosphate buffered saline (PBS). This wash was centrifuged at 1500 rpm/10 min, The pelleted cells were washed two more times in PBS and resuspended at a concentration of 1 x 10^{6}cells/ml.

Nitric Oxide Assay - Nitric oxide content in solutions is determined with the Griess Reagents (0.8% sulfanilic acid and 0.5% N,N,-dimethyl-alpha-naphthylamine, both in 5N acetic acid) (Baxter Medical Corp., Sacramento, CA.) The assay colorimetrically measures nitrite concentration which correlates with Nitric Oxide. In a typical experiment, 10,000 cells/100ul are placed in a well of a 96 well microtiter plate with I00 microliters of each Griess Reagent. The plate is incubated for 10 min at 37 degrees C. Absorbance is read at 546 nm on a BioRad model 450 microplate reader.

RESULTS

Phenotype Experiment - Phenotypic analysis of the lymphocyte gated peripheral blood mononuclear cells of an insulin dependent diabetic patient was made by flow cytometry. A similar analysis was made with blood of a non-diabetic donor. There was higher detectable auto-fluorescence from the diabetic blood. We concluded that this was evidence of non-enzymatic glycosylation of some proteins (AGEs) on or in some of the lymphocytes. The percentage of Natural Killer (NK), B-cells, CD-14+ cells was greater in normal blood. Percentage of T-suppressor cells was lower and T-helper cells was higher in normal blood cells. This data reflects the weaker cellular immunity in the diabetic and the potential of hyperbaric oxygenation in the modulation of immune response.

Serum Nitrite Experiments - Study #1 - This study examined the effects of induction of the shock response with microwaves in rats. Rats were monitored for changes in time to shock and time to death following exposure to 35 GHz. Changes in response of the mitogenic response of rat lymphocytes (table 1) and blood nitrite levels (table 2) were monitored. The production of nitrites by macrophages was also measured. Nitrite concentrations were determined colorimetrically at 570 nm after incubation of the mixture for ten minutes at 37C. Cell proliferation was determined manually with a hemacytometer and all counts are based upon one hundred microliter samples.

Study #2 - This study identified the effects of supertherapeutic hyperbaric oxygenation given until the rat experiences seizures. The average serum nitrite level and the time to seize for rats in five treatment groups is depicted in table 3.

Table 1. PBMC measurements:

Nitrites (mg/L)

Rat #	pre LPS	post LPS	% increase	Treatment
1	5.91	6.89	16	microwave
2	3.16	13.69	333	microwave
3	1.15	1.81	57	control
4	1.48	2.21	49	control
5	7.16	11.79	65	microwave
6	6.31	18.57	194	microwave
7	6.58	6.85	4	control
8	7.22	7.29	1	control
9	29.54	45.27	53	microwave
1 0	5.83	9.12	56	microwave
1 1	11.04	39.56	258	microwave
1 2	11.44	12.62	10	microwave
1 3	2.34	2.41	3	control
1 4	2.66	2.75	3	control
1 5	6.81	4.05	-41	control
1 6	2.01	4.13	1 0 5	control

treatment group	pre LPS		post LPS		% increase	
	avg	sd	avg	sd	avg	sd
microwave	10.05	8.35	19.69	14.51	1 2 3	1 2 2
control	3.78	2.61	3.94	2.10	22	45

pre LPS t-value = 5.36 (pooled sd)
post LPS t-value = 10.93 (pooled sd)
% increase t-value = 21.97 (pooled sd)
t-crit = 3.3257 (two sided, df=14, p<0.005)
t-crit = 2.9768 (two sided, df=14, p<0.01)

There are less than five chances in a thousand that no difference exists between the two treatment groups for pre LPS, post LPS and % increase values. With the assumption that the microwave treatment group has higher values than the control group there are less than five chances in a thousand that the difference is less than 2 for the pre LPS, less than 11 for the post LPS and less than 86% for % increase. Nitrite was measured in four aliquots of each sample and the standard deviation for each sample remained fairly consistent at 1.4 mg/L.

Table 2. Serum Measurements:

Rat #	Nitrites (mg/L)	Treatment
1	10.68	microwave
2	10.05	microwave
3	11.16	control
4	10.33	control
5	0.68	microwave
6	1.35	microwave
7	5.27	control
8	5.61	control
9	5.23	microwave
10	8.98	microwave
11	7.10	microwave
12	4.46	microwave
13	18.98	control
14	20.29	control
15	8.35	control
16	9.08	control

treatment group	average	standard deviation
microwave	6.06	3.80
control	11.14	5.64

t-value = 4.67 (pooled sd)
t-crit = 3.3257 (two sided, df=14, $p < 0.005$)
t-crit = 2.9768 (two sided, df=14, $p < 0.01$)
(treatment1-treatment2)-2 is significant at $p < 0.005$

There are less than five chances in a thousand that no difference exists between the two

treatment groups. With the assumption that the microwave treatment group has lower nitric

oxide in the serum than the control group there are less than five chances in a thousand that the

difference is less than 2.

Table 3. Blood nitrite levels in rats exposed toxic doses of hyperbaric oxygen

	CONTROL	ARG DEF	ARG SUPP	ARG INJ	LNAME
AVERAGE	7.5	2.2	8.3	10.7	20.0
STDEV	5.1	4.6	7.7	14.3	18.5
AVG SEIZE	20 (0)	23 (2)	15 (1)	27 (1)	14.5 (10)
AVERAGE	10.9	13.7	27.1	25.0	49.6
STDEV	9.3	5.5	11.68	33.5	41.0
AVERAGE	10.5	15.8	26.9	24.9	49.1
STDEV	8.4	7.0	12.9	33.5	41.4
AVERAGE	10.1	10.7	21.7	22.0	47.0
STDEV	8.5	5.6	11.7	32.0	40.0

CONCLUSIONS

These studies demonstrate the complexity of the interaction between nitric oxide, oxygen derivatives of nitric oxide and oxygen, and enzymes which modulate the concentrations of these molecules.   The results from these studies indicate the need for further investigation of the enzymes associated with the modulation of oxygen and nitric oxide concentrations and the physiological effects associated with the effects of the interactions between oxygen, nitric oxide and associated enzymes.   In addition, these studies indicate that the functionality of the immune and inflammatory response is altered directly as a result of this complex interaction. These studies demonstrate that functionality of the immune system may not be identified solely through monitoring the expression of leukocyte surface proteins. The overall effect of the response of immune function may not solely be correlated to macrophage production of nitric oxide and the corresponding toxicity to lymphocytes.

Continued studies must be made into the effects of hyperbaric oxygenation on the function of the immune response and inflammatory response. Dose response studies will be able to specifically identify hyperbaric oxygenation regimens which provide these beneficial effects

THE ROLE OF EXPERIENCE IN TRAINING EFFECTIVENESS

Miguel A. Quiñones
Assistant Professor
Psychology Department

Rice University
P.O. Box 1892
Houston, TX 77251

THE ROLE OF EXPERIENCE IN TRAINING EFFECTIVENESS

Miguel A. Quiñones
Assistant Professor
Psychology Department
Rice University

Abstract

For a training program to be effective, trainees must not only master the trained material but also transfer their learning back to the work environment. Research suggests that one factor which can affect a trainee's ability to maintain or improve upon their level of performance acquired during training is the extent to which they get opportunities to performed the trained tasks on the job. More experience with trained tasks is likely to lead to greater skill maintenance and generalization. This study examined the role of experience in training effectiveness by reviewing the existing literature regarding the relationship between experience and performance. In addition, a meta-analysis of the study results was conducted in order to assess the extent to which the results vary as a function of other study characteristics (e.g., measurement of experience, level of analysis, criteria). Finally, implications for training effectiveness research and practice are discussed.

THE ROLE OF EXPERIENCE IN TRAINING EFFECTIVENESS

Miguel A. Quiñones

Introduction

Recent technological and demographic trends suggest that employee training will become an increasingly critical component of an organization's competitive advantage (Goldstein & Gilliam, 1990; Rosow & Zager, 1988). Given these trends, it is important that the training programs employed by organizations have the desired result back in the workplace. Unfortunately, research suggests that only about 10% of training program expenditures result in performance changes on the job (Georgenson, 1982). Part of the problem stems from the lack of adequate program evaluation and the focus on training methods and techniques which employ the most recent "bells and whistles" while ignoring the factors which facilitate or impede the successful implementation of learned skills (Goldstein, 1993; Ford, Quiñones, Sego, & Speer, 1992).

Past research examining the "transfer problem", or lack of adequate implementation of trained skills to the job, have tended to focus on training design issues such as identical elements, task repetition, and training delivery (see Cormier & Hagman, 1987; Quiñones, Ford, Sego, & Smith, in press). Ford, et al. (1992) suggested that a critical, but relatively unexplored, aspect of training effectiveness is the extent to which trainees receive opportunities to perform trained tasks on the job. Their results suggest that there are wide differences in the trained tasks performed by trainees months after leaving the training program. In addition, Ford, et al. (1992) as well as subsequent research (e.g., Quiñones, et al., in press) have identified a number of individual, work context, and organizational factors related to a trainee's level of opportunity to perform trained tasks.

While it seems obvious that trainees need to practice their trained skills if they are to maintain and improve upon their level of performance acquired during training, little research exists which directly addresses this issue. Supporting evidence was presented by Quiñones and Ford (1993) who found a positive relationship between the number of times a trainee performed a trained task in the first eight months on the job, and their level of performance on that task. This relationship existed for over 60% of the tasks examined even after accounting for differences in general cognitive ability and initial level of learning.

Given the importance of obtaining experience with trained tasks to maintain performance levels, the present study further examined the role of experience in job performance. Specifically, the present study sought to (1) summarize the existing literature regarding the role of work experience in affecting job performance, (2) conduct a meta-analysis to determine the effect of work experience on job performance, and (3) discuss implications of the findings for training effectiveness research and practice.

Literature Review

Identification of Relevant Literature

Several criteria were used to select studies for this review. First, the search was limited to studies examining the relationship between work experience and job performance. Second, the search included studies in the published literature as well as military technical reports. Third, only studies which were empirical in nature were included.

Given these search constraint decision rules, a search using Psych-Lit as well as ABI-Inform was conducted. These two databases cover most major journals in the fields of psychology, education, human resource management, and organizational behavior. A manual search of the last two years of key journals in applied psychology and human resource management was also conducted. In addition, a search of the relevant military technical reports was carried out.

The search resulted in the identification of 22 studies containing 36 useful statistics (e.g., correlations) representing a total sample size of 25,799 individuals. Table 1 presents summary information for all studies. Specifically, the sample characteristics, experience measures, criterion measures, and observed correlations are presented.

Insert Table 1 Here

The studies in Table 1 examined the relationship between work experience and job performance for a variety of occupations ranging from skilled laborers (e.g., garment workers, mechanics) to professionals (e.g., managers, clerical). Individuals from public (military, government) and private sector jobs were also sampled in the studies. One study used undergraduate students in a laboratory setting. In terms of criteria, a substantial proportion of the studies used supervisory, self, or peer ratings of performance while others used more objective measures of performance such as hands-on, work sample, or profit measures. Some studies used both types of measures.

Measures of Work Experience

A examination of Table 1 reveals a wide range of measures of work experience. This variability illustrates the lack of consistency in research regarding work experience. Only a few of the studies made any mention of the fact that different measures of experience may be used with perhaps different results (e.g., Hofmann, Jacobs, & Gerras, 1992; Borman et al., 1993). Most measures of experience employed by researchers fell into three main categories.

Time . Most studies employed a time-based measure of experience. These include time spent in a particular job, time in the company, or total time spent in a given occupation. These type of measures are often

referred to as <u>tenure</u>. For example, Giniger, Dispenzieri, and Eisenberg (1983) correlated the amount of time spent in a garment worker position with average hourly piece-rate wages, while McEnrue (1988) correlated the amount of time a manager had spent in their restaurant with the restaurant's profits.

 <u>Amount</u>. A second group of studies measured work experience as an amount. Specifically, these studies defined work experience as the number of times performing a particular task. Individuals performing a task more times are viewed as having more work experience. For example, Spiker, Harper, and Hayes (1983) correlated the number of times Army automotive mechanics performed starter and generator repairs with their scores on a hands-on proficiency test.

 <u>Type</u>. Finally, a few studies categorized an individual's work experience into types. For example, Laxar and Olson (1978) compared the job performance levels of instructors in a Navy training course with those of recent graduates. Instructors were presumed to have more work experience than recent graduates both in a qualitative and quantitative sense. Pinder and Schroeder (1987) defined experience as the degree of similarity between a person's previous job and their current one. The more similar the previous job is to the current one, the more work experience the person is presumed to bring to the current job.

Levels of Analysis

 A second dimension which characterizes the differences in the measurement of work experience is the level of analysis (see Klein, Dansereau, & Hall, 1994; Rousseau, 1985). Most of the literature of work experience has focused on the individual as the unit of analysis. However, within an individual, studies have examined task, job, or organizational work experience. While work experience was not measured at each of these three levels, each of the dimensions described above (time, amount, type) can be conceptualized and measured at each of these levels of analysis.

 <u>Task</u>. Individuals can vary in their level of experience performing specific tasks (Ford et al., 1992). First, they can perform a particular task a given number of times (amount). Second, individuals can vary in the type of task that they have performed (type). Some may perform simple tasks whereas others may perform more difficult, complex, and critical tasks. Finally, individuals can vary in the amount of time spent working on a given task (time).

 It is clear that each measure of task-level experience captures a somewhat unique portion of an individual's overall level of experience. For example, two people can perform any task the same number of times but differ in the difficulty or criticality of the task. Similarly, individuals may spend the same amount of time performing a task but differ in the number of times they perform the task in that time span.

 <u>Job Experience</u>. It is also possible to measure a person's experience at the job level of analysis. First, individuals can differ in the total number of jobs that they have held (amount). Sometimes task level experience is aggregated to the job level (e.g., Lance, Hedge, & Alley, 1987). This corresponds to the breadth measure described by Ford, et al. (1992). Individuals also can have different experiences by performing different types of jobs which vary in terms of prestige, difficulty, criticality (type). Finally, differences in work experience can be represented by the amount of time spent in a particular job, or job tenure (time).

<div align="center">34-5</div>

Organizational Experience. Differences in experience can also exist at the organizational level of analysis. First, individuals can vary in the number of organizations in which they have worked (amount). Second, organizational experience can vary depending on the type of organization in which a person has worked such as manufacturing, research and development, etc. (type). Finally, organizational experience can vary depending on the amount of time spent in a given organization (time).

Meta-Analysis

The two dimensions described above serve as a useful categorization scheme for examining the relationship between work experience and job performance in a more systematic manner. A meta-analysis was conducted on the 22 studies (36 correlations) described in Table 1 using the three measures of experience (amount, time, type) and level of analysis (task, job, organization) as potential moderators. The first step in this process is to examine the relationship between work experience and job performance as a function of the type of experience measure used. This involved coding the articles with respect to the characteristics of interest and examining the relationship between experience and performance across all studies in a quantitative manner.

Method

Coding of Studies

Initially, ten studies were coded by three independent coders and results were discussed to facilitate reliability and consistency in coding. All studies were then coded by two independent coders using the experience dimensions described above. In addition to experience and level of analysis, the type of criteria used was also coded. These were divided into two categories; hard and soft criteria. Hard criteria represent fairly objective measures of performance such as production units, amount of sales, job samples, or work simulations. Soft criteria were supervisory, peer, and self ratings of job performance. The coders showed almost perfect agreement in coding experience and performance measures.

One assumption that is made when conducting a meta-analysis is that the statistics used in the calculations are independent. That is to say, they come from different samples (Hunter & Schmidt, 1990). A few of the studies in the database contained more than one statistic for each sample. These statistics were combined (averaged) only when they reflected similar study characteristics such as experience measure, job performance measure, and level of analysis. If the statistics reflected different characteristics, they were analyzed separately. This situation occurred in only 2 out of the total 22 studies. It was believed that the independence assumption was not severely violated (Hunter & Schmidt, 1990).

Meta-Analytic Procedures

Meta-analytic procedures were used to examine the overall effect of experience on performance, as well as the potential moderators of this relationship (Hunter & Schmidt, 1990; Raju, Burke, Normand, & Langlois, 1991). Meta-analysis is a statistical technique which allows for the aggregation of the results across studies and correct for various statistical artifacts in order to get an estimate of the true relationship between two variables in the population.

For purposes of this meta-analysis, all study statistics which reflect the relationship between experience and performance were converted into correlations (see Hunter & Schmidt, 1990). Correlations are easily manipulated and provide a standardized measure of the strength and direction of a relationship which is easily interpretable. Meta-analytic procedures require that each observed correlation be weighted by the sample size in order to calculate a mean weighted correlation (mean r) across all studies examined. The standard deviation of this mean r (SD_r) is then computed which reflects the variability in the relationship between work experience and job performance across studies.

The total variation across studies is composed of several key elements. These include, true variation in the population, variation due to sampling error, and variation due to other statistical artifacts such as reliability and range restriction. By accounting for variation due to statistical artifacts, one can obtain a better measure of the true variability around the population correlation. Because of limited reliability information, and the fact that some studies made corrections for reliability when reporting the results, only corrections due to sampling error were made (e.g., bare-bones Meta-Analysis). In this situation, the mean r is also the estimate of the population correlation rho (ρ).

If a large portion of the variability in the observed correlations across studies can be accounted for by these statistical artifacts, one can conclude that the relationship between work experience and job performance is fairly constant (Ones, Viswesvaran, & Schmidt, 1993). In addition, if the 95% confidence interval around the mean correlation does not include zero, it can be concluded that the relationship between work experience and job performance in the population is non-zero (Finkelstein, Burke, & Raju, 1994; Whitener, 1990).

In addition to estimating the population correlation, meta-analysis allows one to determine the extent to which the observed relationship between work experience and job performance depends on other factors. If the results show that after accounting for statistical artifacts, a substantial amount of variability in the correlations across studies remains, other factors may be used to help account for this variability. In this study, three classes of moderators were examined. These included type of performance measure (soft vs. hard), experience measure (amount, type, & time), and level of analysis (task, job, and organization).

In conducting moderator analyses, separate meta-analysis calculations were computed for each subset of studies (e.g. hard and soft criteria). Two pieces of information were used to examine the presence of a moderated relationship. First, the variation around the mean correlation for each subset was examined. If this variation was close to zero, it was concluded that the characteristic used to divide up the studies accounted for the initial variation in correlations across studies. Second, a confidence interval around the mean correlation was constructed for each subset of studies based on the standard error of the estimated population correlation rho (Finkelstein, Burke, & Raju,

1994; Whitener, 1990). A lack of overlap between confidence intervals led to the conclusion that the overall effect of experience on performance differed by the characteristic used to divide up the studies.

Results

Table 2 presents the results of the meta-analyses conducted on the entire database as well as the results of the moderator analyses. Each is discussed in turn.

Insert Table 2 Here

The first set of analyses was conducted using all 36 available correlations with a total sample size of 25,799. The mean weighted r was .23 (SD_r = .110). The 95% confidence interval around the mean correlation did not include zero, suggesting that relationship between experience and performance in the population is positive. Only 10.23% of the variance in the observed correlations was accounted for by sampling error suggesting that other study characteristics may moderate the relationship between experience and performance.

The second set of analyses involved dividing up the studies according to type of criteria used, experience measure, and level of analysis. Of the 36 correlations examined, exactly half of the studies (K = 18 each) used soft (N = 21,600) and hard (N = 4,199) performance criteria. The results indicate that experience had a stronger relationship with hard (mean r = .35, SD_r = .155) than with soft (mean r = .20, SD_r = .071) performance criteria. Both sets of confidence intervals did not include zero, suggesting a positive relationship between experience and both types of performance measures. The 95% confidence intervals did not overlap suggesting that the strength of the relationship is different for the two criterion measures. A substantial amount of unexplained variance remained for soft and hard criterion measures (14.12% and 13.77% variance explained, respectively).

The next potential moderator examined was the type of experience measure used. The results show that the strongest relationship between experience and performance occurs when amount of experience (mean r = .25, SD_r = .179) was used. This was followed by time (mean r = .23, SD_r = .105) and finally type (mean r = .18, SD_r = .032). The confidence intervals around these three values revealed that they are all positive. However, a substantial amount of overlap existed across the three credibility intervals. The amount of variance explained increased only for studies using type of experience (100% variance explained). For the other two subsets, a substantial amount of unexplained variability remained (13.54% and 8.80% explained).

The final potential moderator examined was the level of analysis examined by each study. The results indicate that the largest relationship between experience and performance occurred when experience was measured at the task level (mean r = .33, SD_r = .016) when compared to those studies examining this relationship at the job

(mean r = .21, SD_r = .095) or organizational level of analysis (mean r = .13, SD_r = .105). The confidence intervals around these estimates indicate that they were all positive. The confidence intervals suggest that the relationship between task-level experience is significantly greater than job- and organizational-level experience.

Discussion

The purpose of this study was to review the literature examining the relationship between work experience and job performance to further understand the role of experience in training effectiveness. Two dimensions of work experience were developed using the three measures of experience (amount, time, and type) and three levels of analysis (task, job, and organization). In addition, the effects of experience on two types of performance measures (soft vs. hard) were examined.

Results of the literature review revealed that most researchers used a time-based measure of experience (e.g., tenure). A few measured work experience as an amount and still fewer measured the type of work experience. In addition, most studies used job level measures of work experience. However, some studies examined task level and organizational level experience. It is clear from this review that individuals are using the same term (e.g., experience) to refer to very different measures.

The results of the meta-analyses revealed that the relationship between experience and performance was positive regardless of the type of measure used. The relationship was stronger when hard performance measures such as work samples were used as compared to soft performance measures such as supervisory ratings. The meta-analyses results also revealed some variation in the relationship between work experience and job performance as a function of the type of experience measure used. The strongest relationship occurred between amount of experience and performance while the weakest involved type of experience.

Finally, some variability in the relationship between experience and performance as a function of level of analysis was found. Task level experience had the strongest relationship with performance whereas organizational level showed the weakest. The results of these analyses should be interpreted with caution as only two correlations were based on organizational level work experience.

Conclusions and Future Research Directions

The results of this study suggest a number of conclusions and future research directions regarding the role of the work experience in training effectiveness. First, it clear that more experience leads to better performance. However, it also appears that task repetition (amount of experience at the task level) is most strongly related to performance. This suggests that trainees must be able to get very specific experience with the trained material if they are to maintain a high level of performance.

However, it is still not clear how much experience is needed to maintain performance at a specified level. This amount is likely to vary as a function of task characteristics such as difficulty, complexity, or criticality, as well

as individual characteristics such as general cognitive ability and self-efficacy. More research is needed to address these issues.

Given the importance of experience, it is necessary to examine the factors which affect the types of assignments trainees receive. Past research suggests that supervisors are a key gatekeeper in determining the types of assignments received by recent trainees (Quiñones, et al, in press). Supervisors must be more aware of their assignment patterns if they are to ensure that everyone maintains their level of performance over time. Coworkers have also been found to affect the opportunity to perform trained tasks (Ford, et al., 1992). In addition, situational constraints such as the availability of tools and equipment is also likely to influence a trainees ability to practice their skills.

The comparisons in this study made regarding the various experience measures represent between-subject comparisons. Future research should examine the relationship between various measures of work experience and job performance within an individual. For example, Ford, Sego, and Teachout (1991) found a positive relationship between the number of times performing a task and a hands-on measure of task experience after accounting for differences in job tenure. Similar results were found by Quiñones and Ford (1993) as well as Dubois and McKee (1994). Thus, the results presented here may not necessarily capture the unique effects that various measures of work experience can have on job performance.

Future research could also examine the relationship between various measures of work experience and other organizational outcomes, in addition to job performance. For example, organizational tenure is more likely to be related to organizational commitment than is task level experience. Conversely, task level experience is more likely to be related to self-efficacy than is organizational level experience. Experience is also relevant for the study of expertise and the development of expert mental models (Cooke & Schvaneveldt, 1988; Howell & Cooke, 1989). It is important to know whether expertise develops as a result of time or as a function of the number of times a particular task is performed.

REFERENCES

Barrick, M.R. & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-26.

Borman, W.C., Hanson, M.A., Oppler, S.H., Pulakos, E.D., & White, L.A. (1993). Role of early supervisory experience in supervisor performance. Journal of Applied Psychology, 78, 443-449.

Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt, W.C. Borman, & Associates (eds.), Personnel Selection in Organizations. San Francisco, CA: Jossey-Bass.

Cobb, B.B. (1968). Relationship among chronological age, length of experience, and job performance ratings of air route traffic control specialists. Aerospace Medicine, 39, 119-124.

Cooke, N.M. & Schvaneveldt, R.W. (1988). Effects of computer programming experience on network representations of abstarct programming concepts. International Journal of Man-Machine Studies, 29, 407-427.

Cormier, S.M. & Hagman, J.D. (1987). Transfer of Learning. San Diego: Academic Press.

DuBois, D. & McKee, A.S. (1994, May). Facets of work experience. Paper presented at the 9th Annual Conference of the Society for Industrial and Organizational Psychology, Nashville, TN.

Early, P.C., Lee, C., & Hanson, L.A. (1990). Joint moderating effects of job experience and task component complexity: Relations among goal setting, task strategies, and performance. Journal of Organizational Behavior, 11, 3-15.

Finkelstein, L.M., Burke, M.J., & Raju, N. (1994, April). Age discrimination in employment: An integrative analysis. Paper presented at the 9th Annual Meeting of the Society for Industrial and Organizational Psychology, Nashville, TN.

Ford, J.K., Kraiger, K., & Schechtman, S.L. (1986). Study of race effects in objective indices and subjective evaluations of performance: A meta-analysis of performance criteria. Psychological Bulletin, 99, 330-337.

Ford, J.K., Quiñones, M.A., Sego, D.J., & Sorra, J. (1992). Factors affecting the opportunity to perform trained tasks on the job. Personnel Psychology, 45, 511-527.

Ford, J.K., Sego, D.J., Quiñones, M.A., & Speer, J. (1991, April). The construct of experience: A review of the literature and needed research directions. Paper presented at the Sixth Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Ford, J.K., Sego, D.J., & Teachout, M.S. (1991, April). The effects of cognitive ability, job tenure, and amount of task experience on task proficiency. Paper presented at the 6th Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Georgenson, D.L. (1982). The problem of transfer calls for partnership. Training and Development Journal, 36, 75-78.

Giniger, S., Dispenzieri, A., & Eisenberg, J. (1983). Age, experience, and performance on speed and skill jobs in an applied setting. Journal of Applied Psychology, 68, 469-475.

Goldstein, I.L. (1993). Training in organization: Needs assessment, development, and evaluation (3rd ed). Monterey, CA: Brooks/Cole.

Goldstein, I.L. & Gilliam, P. (1990). Training systems in the year 2000. American Psychologist, 45, 134-143.

Gordon, M.E. & Fitzgibbons, W.J. (1982). Empirical test of the validity of seniority as a factor in staffing decisions. Journal of Applied Psychology, 67, 311-319.

Gould, S. & Hawkins, B.L. (1978). Organizational career stage as a moderator of the satisfaction-performance relationship. Academy of Management Journal, 21, 434-450.

Hall, D.T. & Mansfield, R. (1975). Relationships of age and seniority with career variables of engineers and scientists. Journal of Applied Psychology, 60, 201-210.

Hofmann, D.A., Jacobs, R., & Gerras, S.J. (1992). Mapping individual performance over time. Journal of Applied Psychology, 77, 185-195.

Hogan, J.C. (1991). Physical abilities. In M.D. Dunnette & L.M. Hough (eds.), Handbook of Industrial and Organizational Psychology, (Vol 2), pp. 753-831. Palo Alto, CA: Consulting Psychologists Press.

Howell, W.C. & Cooke, N.J. (1989). Training the human information processor: A review of cognitive models. In I.L. Goldstein & Associates (eds.), Training and Development in Organizations, San Francisco, CA: Jossey-Bass.

Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

Hunter, J. & Schmidt, F. (1990). Methods of Meta-Analysis. Newbury Park, CA: Sage.

Jacobs, R. & Hofmann, D.A. (1990). Performance and seniority. Human Performance, 3, 107-121.

Katz, R. (1978). The influence of job longevity on employee reactions to task characteristics. Human Relations, 31, 703-725.

Klein, K.J., Dansereau, F., & Hall, R.J. (1994). Levels issues in theory development, data collection, and analysis. Academy of Management Review, 19, 195-229.

Lance, C.E., Hedge, J.W., & Alley, W.E. (1987). Ability, experience, and task difficulty predictors of task performance. (AFHRL-TP-14). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Laxar, K. & Olson, G.M. (1978). Human information processing in navigational displays. Journal of Applied Psychology, 63, 734-740.

Maranto, C.L. & Rodgers, R.C. (1984). Does work experience increase productivity? A test of the on-the-job training hypothesis. The Journal of Human Resources, 19, 341-357.

McDaniel, M.A., Schmidt, F.L., & Hunter, J.E. (1988). Job experience correlates of job performance. Journal of Applied Psychology, 73, 327-330.

McEnrue, M.P. (1988). Length of experience and the performance of managers in the establishment phase of their careers. Academy of Management Journal, 31, 175-185.

Ones, D.S., Viswesvaran, C., & Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. Journal of Applied Psychology, 78, 679-703.

Petty, M.M. (1974). A multivariate analysis of the effects of experience and training upon performance in a leaderless group discussion. Personnel Psychology, 27, 271-282.

Pinder, C.C. & Schroeder, K.G. (1987). Time to proficiency following job transfers. Academy of Management Journal, 30, 336-353.

Potter, E.H. & Fiedler, F.E. (1981). The utilization of staff member intelligence and experience under high and low stress. Academy of Management Journal, 24, 361-376.

Quiñones, M.A. & Ford, J.K. (1993, May). Individual and task experience effects on transfer of training: A field study. Paper presented at the 8th Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Quiñones, M.A., Ford, J.K., Sego, D.J., & Smith, E.M. (in press). The effects of individual and transfer environment characteristics on the opportunity to perform trained tasks. Training Research Journal.

Raju, N.S., Burke, M.J., Normand, J., & Langlois, G.M. (1991). A new meta-analytic approach. Journal of Applied Psychology, 76, 432-446.

Rakestraw, T.L. & Weiss, H.M. (1982). The interaction of social influences and task experience on goals, performance, and performance satisfaction. Organizational Behavior and Human Performance, 27, 326-344.

Ree, M.J. & Earles, J.A. (1992). Intelligence is the best predictor of job performance. Current Directions on Psychological Science, 1, 86-89.

Rosow, J.M. & Zager, R. (1988). Training - The competitive edge. San Francisco: Jossey-Bass.

Rowe, P.M. (1988). The nature of work experience. Canadian Psychology, 29, 109-115.

Rousseau, D.M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. In L.L. Cummings and B.M. Staw (eds.), Research in Organizational Behavior, 75, 1-37. Greenwich, CT: JAI Press.

Schmidt, F.L., Hunter, J.E., & Outerbridge, A.N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. Journal of Applied Psychology, 71, 432-439.

Schmidt, F.L., Hunter, J.E., Outerbridge, A.N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. Journal of Applied Psychology, 73, 46-57.

Schmitt, N. & Cohen, S. (1989). Internal analyses of task ratings by job incumbents. Journal of Applied Psychology, 74, 96-104.

Schmitt, N., Gooding, R.Z., Noe, R.A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Schwab, D.P. & Heneman, H.G. (1977). Effects of age and experience on productivity. Industrial Gerontology, Spring, 113-117.

Spiker, V.A., Harper, W.R., & Hayes, J.F. (1985). The effects of job experience on the maintenance proficiency of army automotive mechanics. Human Factors, 27, 301-311.

Teachout, M.S. (1991, April). Understanding the work experience construct in personnel research and practice. Symposium conducted at the Sixth Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Whitener, E.M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. Journal of Applied Psychology, 75, 315-321.

Table 1

Summary Table of Experience Literature

Study	N	Sample Description	Experience Measure	Level	Criteria Measure	Level	r
Borman et al. (1993)	570	Army Personnel	time in supervisory job	job	supervisor ratings	job	.14
Cobb (1968)	568	air traffic controllers	time since entering duty	job	supervisor ratings	job	-.05
			time since entering duty	job	peer ratings	job	-.02
Earley, Lee, & Hanson (1990)	347	various occupations	time in present job	job	supervisor ratings	job	.10
Giniger, Dispenzieri, & Eisenberg (1983)	212	garment workers (speed jobs)	job tenure	job	average hourly piece-rate wage	job	.47
	455	garment workers (skill jobs)	job tenure	job	average hourly piece-rate wage	job	.46
Gordon & Fitzgibbons (1982)	162	sewing machine operators	tenure with company	org	quarterly efficiency average	job	.23
			time on job	job	quarterly efficiency average	job	.15
			interjob similarity	job	quarterly efficiency average	job	.20

34-15

Table 1 (cont'd)

Study	N	Sample Description	Experience Measure	Experience Level	Criteria Measure	Criteria Level	r
Gould & Hawkins (1978)	132	managerial and clerical public employees	time in company	org	supervisor ratings	job	.37
Hall & Mansfield (1975)	290	professionals in R&D companies	time in company	org	self-ratings of performance	job	-.11
	90	professionals in R&D companies	time in company	org	self-ratings of performance	job	.01
Jacobs, Hofman, & Kriska (1990)	126	firefighters	job tenure	job	supervisor ratings	job	.22
			job tenure	job	physical ability work sample	job	.27
Katz (1978)	89	public sector employees	time in present job	job	supervisory ratings of overall job performance	job	-.02
Lance, Hedge, & Alley (1987)	217	Air Force jet engine mechanics	job experience composite	job	walk-through performance test	task	.07
	217	Air Force jet engine mechanics	job experience composite	job	overall WTPT performance ratings	task	.02
	217	Air Force jet engine mechanics	job experience composite	job	task performance self-ratings	task	.22
	217	Air Force jet engine mechanics	job experience composite	job	task performance supervisor ratings	task	.18
	217	Air Force jet engine mechanics	task experience composite	task	walk-through performance test	task	.20

34-16

Table 1 (cont'd)

Study	N	Sample Description	Experience Measure	Experience Level	Criteria Measure	Criteria Level	r
Lance, Hedge, & Alley (1987)	217	Air Force jet engine mechanics	task experience composite	task	overall WTPT performance ratings	task	.08
	217	Air Force jet engine mechanics	task experience composite	task	task performance self-ratings	task	.54
	217	Air Force jet engine mechanics	task experience composite	task	task performance supervisor ratings	task	.23
Laxar & Olson (1978)	14	Navy submarine officers	instructors vs. recent graduates	job	errors in performing line-of-sight computer task	task	.54
Maranto & Rodgers (1984)	191	wage-hour investigators	time with division	org	wages recovered for claimant	job	.21
			years of previous work experience	job	wages recovered for claimant	job	.15
McDaniel, Schmidt, & Hunter (1988)	16,058	various occupations	tenure in occupation	job	supervisor ratings	job	.21
McEnrue (1988)	89	restaurant managers	tenure in current restaurant	job	restaurant profits	org.	.19
			tenure with org.	org.	restaurant profits	org.	.30
Petty (1974)	100	ROTC students	experience in leaderless group discussion (manipulated)	task	peer ratings	task	.17

Table 1 (cont'd)

Study	N	Sample Description	Experience Measure	Experience Level	Criteria Measure	Criteria Level	r
Pinder & Schroeder (1987)	354	managers	change in job function	job	time to proficiency self-ratings	job	.13
			perceived job similarity	job	time to proficiency self-ratings	job	.21
			number of previous transfers	job	time to proficiency self-ratings	job	.09
Potter & Fiedler (1981)	130	Coast Guard	months in service	org	supervisor ratings	job	.18
Rakestraw & Weiss (1981)	174	college students	experience with card sorting task (manipulated)	task	# cards sorted in 5 min.	task	.25
Schmidt, Hunter & Outerbridge (1986)	1474	various occupations	job tenure	job	job knowledge	job	.53
			job tenure	job	work sample	job	.47
			job tenure	job	supervisor ratings	job	.33
Schwab & Heneman (1977)	124	assembly workers	tenure with company	org	average hourly productivity	job	.42

Table 1 (cont'd)

Study	N	Sample Description	Experience Measure	Experience Level	Criteria Measure	Criteria Level	r
Spiker, Harper, & Hayes (1983)	70	Army automotive mechanics	self-reported number of starter repairs performed	task	overall profiency (hands on)	task	.69
			self-reported number of starter repairs performed	task	repair time (hands on)	task	.69
			self-reported number of starter repairs performed	task	overall accuracy (hands on)	task	.57
			self-reported number of generator repairs performed	task	overall profiency (hands on)	task	.73
			self-reported number of generator repairs performed	task	repair time (hands on)	task	.58
			self-reported number of generator repairs performed	task	overall accuracy (hands on)	task	.61

34-19

Table 2

Meta-Analysis Results

Analysis	N	K	mean r	SDr	SEρ	% Acc.	95% CI
Overall	25,799	36	.23	.110	.006	10.23	.21 - .24
Performance Measure							
Soft	21,600	18	.20	.071	.006	14.12	.19 - .21
Hard	4,199	18	.35	.155	.013	13.77	.33 - .38
Experience							
Amount	1,202	7	.25	.179	.026	13.54	.20 - .30
Time	23,727	26	.23	.105	.006	8.80	.21 - .24
Type	870	3	.18	.032	.033	100.0	.11 - .24
Level of Analysis							
Task	3,071	9	.33	.138	.016	11.34	.29 - .36
Job	22,385	25	.21	.095	.006	11.05	.20 - .23
Organization	343	2	.13	.105	.053	48.37	.03 - .23

NOTE: N = total sample size; K = number of correlations; mean r = mean weighted correlation across studies; SD_r = standard deviation of mean r; SE_ρ = standard error of population correlation; % Acc. = percent of variance accounted for by statistical artifacts; 95% CI = 95 percent confidence interval around mean r.

AETMS: Analysis, Design and Development

R. Ramesh

Associate Professor

Department of Management Science & Systems

School of Management

State University of New York at Buffalo

Buffalo, NY 14260.

1 Introduction

The development and maintenance of all aircrews and support personnel in a state of mission-readiness is a primary goal of the Air Force. The training of all air staff is systematically carried out at both formal training schools and the field units of the Air Force to meet this objective. The formal training broadly consists of primary air education for all aircrew and support personnel, undergraduate aircrew training, graduate level combat crew training, technical training for the maintenance personnel and continuation training for all air staff. In 1993, all formal training programs have been consolidated and integrated under the Air Education and Training Command (AETC) of the Air Force. AETC is headquartered at Randolph AFB and its training programs are broadly organized into the Air University, 19th Air Force and 2nd Air Force. The Air University is responsible for primary air education and is headquartered at Maxwell AFB. The 19th Air Force operates from Randolph AFB and is responsible for all aircrew training. The 2nd Air Force operates from Keesler AFB and is responsible for all technical training. The training wings under each of these units are located at various bases and conduct both initial and continuation training.

Air staff training can be broadly categorized into five types: *flight, static, ground, simulator* and *academic.* Each training program comprises of a set of courses, and each course is a package of programmed instruction modules involving these training types. Training programs entail considerable human and equipment resources, extensive planning, co-ordination and management. The current trend within the Air Force is to design training programs as *total integrated systems* in order to achieve the training objectives at a desired level of cost-effectiveness. The Training Management Systems (TMS) are intended to be the *system integrators* by providing the required connectivity among the users, distributed database management, information support and decision support at various levels of the organization.

Several forms of TMS at different levels of sophistication and implementation currently exist in the Air Force. Most of the fully functional systems are in-house developed several years back. These systems are virtually standalone, partially interoperable within a wing and provide limited process automation and basic functional support. However, the growth in the magnitude and complexity of the training programs strongly underscores the need for greater process automation, efficient information/decision support and a commandwide integration of all training management systems. The concept of Air Education Training management System (AETMS) is a product of this need, and has evolved through several stages of analysis and refinement at AETC. Beginning with a vision of a single comprehensive system supporting all user requirements commandwide, this concept has evolved to a fully interoperable suite-of-systems under an unified umbrella. This concept can be further refined by regarding AETMS as a commandwide system platform consisting of standardized hardware and software environments housing a variety of information/decision support systems for training management. The system platform specifies the hardware/software requirements of any

system that is developed under the umbrella. This definition provides the *flexibility* to develop any new system under the umbrella and ensures its *interoperability* with the other systems on the platform. The absence of such backbone specifications leads to standalone systems and a host of problems with their integration. Further, the platform concept would greatly facilitate all the tasks in system acquisition and minimize the possible errors.

The objectives of this study can be summarized as follows: (i) develop a model of the AETMS platform in a client-server distributed LAN environment, (ii) develop a set of functional and technical specifications for systems on the platform, (iii) specialize the model to the suite-of-systems at AETC, (iv) identify and delineate the short-run developmental tasks, and (v) program the short-run tasks with a focus on the long-term vision. The detailed results of this study are presented in the comprehensive report submitted to AL/HRA and AETC. We present the design model for the AETMS platform in this report.

2   AETMS Platform: A Design Model

The International Standards Organization (ISO) presents a seven layer model of Open System Integration (OSI). The seven layers are: *physical, data link, network, transport, session, presentation* and *application*. We employ the OSI model as a basis for the design of the AETMS platform. The first four layers specify the requirements of the underlying communication and computing hardware environment. The fifth layer specifies the requirements of inter-process communication sessions. The sixth layer specifies data representation and the session interface requirements. The seventh layer addresses the residence and user interface requirements of the application systems and databases. We model the AETMS platform by combining these seven layers into three broad layers of the overall system architecture. These layers are: distributed application layer, network layer and middleware layer. The distributed application layer provides the architectural requirements for the distributed database environments supported in the platform. The network layer provides the architecture of the communication and computing environment in which the platform is set up. The middleware layer integrates these two layers by providing the protocols for a co-ordinated communication and management of tasks among the client-server stations in the network. From the point of view of the OSI model, the distributed application layer represents the OSI layer 7, the middleware layer represents layers 6 and 5, and the network layer represents the layers 4 - 1.

2.1   Distributed Application Layer

At the outset, we assume that all databases on the platform are relational and SQL driven. This is a minimal level of database standardization on the platform necessary to achieve cost-effective interoperability among the systems. A *database* is a collection of *relations* or *data files* managed and

maintained by a database management system (DBMS). A DBMS has an associated *environment* within which all the systems managed by it operate. The databases and related software programs for a specific management function constitute an *application module*. A set of related application modules are combined into an *application system*, and the set of application systems managed by a DBMS run within its environment. For example, consider a TMS at a training wing. The TMS is maintained by a DBMS, and its application systems are curriculum management, configuration management, scheduling management, resource management and so on. The application modules within the scheduling management system for example, are flight scheduling, academic scheduling, simulator scheduling and so on. The modules within a system are functionally related and a system possesses an unique functional identity of its own. Using these basic concepts, we describe the database structure of the AETMS platform as follows.

The AETMS platform contains a set of autonomous but integrated DBMS environments. For example, ATS, JPATS/TIMS, CCT/TMS, ED/TMS and MIS for AETC/HQ represent autonomous DBMS environments that are integrated on the AETMS platform. We term these environments as local. All the local environments are contained within a common, federated environment of the AETMS platform. The federated environment is termed as global. Thus, the structure of data and programs on the AETMS platform is a hierarchy as follows: a set of application-specific *databases* and *programs* constitute an *application module*, a set of functionally-related modules constitute an *application system*, a set of application systems with a unique functional identity for each system operating under a DBMS constitute a *local environment*, and the set of all local environments constitutes the *global environment*.

The object structure and the schematics of the AETMS platform follow logically from the data/program hierarchy as follows. The schema of an application module treats its database relations and programs as *objects entities* and describes them fully. The schema of an application system represents a level of consolidation, treating its modules as objects. The schema of a system describes its modular aggregates and the common data dictionary elements among its modules. The schema of a local environment treats its systems as objects, and describes its system aggregates and the common elements among its systems. Finally, the global schema treats the local environments as objects and provides an aggregate total view of the platform. The schemata of the global environment, local environments, systems and modules provide the *interoperability hooks* among the local environments, systems, modules and data/program structures within a module, respectively. The functional specifications for the distributed application layer are as follows.

1. DBMS Standards

All DBMS environments should be COTS and conform to ANSI-SQL X3-135. A DBMS should be capable of operating within the existing hardware/software environment and adaptable to the potential system expansions/upgrades in the future. A DBMS should be free of proprietary

hooks and any singular dependence on devices such as a mainframes for instance.

2. Local Autonomy

All databases in a local environment should be *locally* owned and managed. Security, integrity, storage representation and application programs should be locally managed by the set of servers/clients hosting the local environment.

3. Data Fragmentation

All forms of database fragmentation should be possible in each local environment. This includes: (i) distributing database relations across several servers, (ii) storing replications/extracted replications/snapshots at both clients and servers, and (iii) storing vertical/horizontal partitions of any relation at clients and servers. Under all forms of fragmentation, the local DBMS should be capable of maintaining both the security and integrity of its databases.

3. Distribution Transparency

The distribution of the data fragments should be invisible to the users. The users should in no way be concerned with how the data is distributed in formulating their queries and updates. Assembling the data required for any of these tasks should be done exclusively by the DBMSs on the platform.

4. Hardware Heterogeneity

The DBMSs managing the local and the global environments on the platform should be capable of operating in a heterogeneous hardware environment comprising of various hardware platforms, network architectures, operating systems and transport protocols. This implies that the distributed application layer should adapt to the network layer of the AETMS platform.

5. Distributed Query Processing

User query processing should be carried out in a distributed fashion to optimize system performance. A distributed processing plan should be determined for each query as follows: (i) assess the current traffic loads in the network, (ii) determine the servers/clients holding the data fragments needed for the query, (iii) develop a set of potential processing plans from (i) and (ii), where each plan specifies the data accessed, processed and transmitted by the holding stations to an assembly point from which the results are delivered to the user, and (iv) choose the processing plan that is most likely to minimize the query response time. In any processing plan, appropriate local processing should be carried out before moving data out. The local environments should incorporate *distributed query optimization models* for this purpose. The query optimization models work in tandem with the platform middleware, and these links are modelled in a latter section.

6. Distributed Integrity/Security Management

The management of semantic integrity constraints is fairly standard. The COTS DBMSs provide facilities for defining these constraints within the application programs. However, constraints of consistency among the fragments of a database are important in the distributed context. The DBMSs managing the environments should provide tools for defining and enforcing these constraints. The proposed hierarchical structure of the schemata can be used to design integrity management protocols for the distributed database architecture on the AETMS platform. Similarly, all the environments in the platform should include security protocols for system-level and user-level access and update of the databases.

7. Distributed Transaction Management

A transaction is a unit of work or concurrency or recovery. All distributed transactions must be atomic, serializable and stable. If any site fails during a distributed transaction such as replica update for instance, all participating subtransactions must be recovered using journal backups. Distributed transaction management becomes critical when fragmentation is high.

8. Distributed Concurrency Control

Concurrency control is necessary to prevent several users from interfering with each other in database access. In a distributed architecture, concurrency control within a transaction becomes important if several processors are simultaneously used in processing the transaction. Serializing parallel transactions and serializing processors within a transaction are necessary for correctness of a transaction. The environments should provide tools for concurrency control such as transaction locking, transaction/processor deadlock detection and avoidance, and replica update synchronization.

2.2 Network Layer

The basis of the AETMS platform is the commandwide communication network. The physical architecture of the *host* is shown in Figure 1. The wide area networks DDN and AUTODIN form the backbone of the DoDwide communication system. The communication network hosting the platform is divided into a set of *segments*, where each segment pertains to an AETC location such as HQ or its wings. The agencies outside AETC such as other MAJCOMs, Navy, and other DoD agencies could access the platform through the wide area networks.

The physical architecture of the AETC/HQ segment is shown in some detail in Figure 1, for purposes of illustration. Similar communication networks should exist at the training wings as well. The AETC/HQ segment consists of four levels of networking: a fully routed FDDI backbone, a network of active FOIRL building hubs that link to a system of passive hubs as well as directly to LANs, a system of passive hubs that link to a set of LANs, and the system of LANs reaching

out to the end devices. The access to the backbone is through fiber, and both fiber and coax links exist from the hubs to the LANs. Both bus and ring topologies are used in the LANs and all types of connection media exist. While most buildings are LAN connected, isolated peer-to-peer chains using Microsoft post offices and modem links to the backbone also exist. The passive hubs are slowly being phased out, and the state-of-the-art modular concentrators are being considered for future expansions and upgrades. We assume full connectivity among all devices in a segment in the following analysis. The segments in the physical architecture are termed *physical segments.*

The assumption of full local connectivity enables a concise logical view of a segment, avoiding the cumbersome technical details of internal physical routings in the network. A physical segment is viewed as a set of fully interconnected LANs with gateway access to the wide area networks, leading to a logical host architecture shown in Figure 2. The segments in the logical architecture are termed *logical segments.* This leads to the following hierarchy of network structures: each user workstation is a *client* and any other device is a *client* or a *server.* A server and a set of clients make up a LAN, a set of interoperable LANs makes up a *segment,* and the set of segments constitutes the *communication network.* The functional specifications for the network layer are fairly standard, and are as follows: *distributed database connectivity, router-to-router protocols, comprehensive distributed computing/management, network management, internetworking transport* and *user interfaces.*

2.3   Middleware Layer

Middleware provides the transparency between the distributed application layer and the network layer. From a functional standpoint, middleware allows distributed applications to communicate across the commandwide network in a transparent manner. This transparency applies to the hardware platform, operating system and the networking protocols installed in any LAN system. Middleware ensures interoperability among a set of distributed stations and database environments regardless of the underlying hardware and software configurations. As a result, middleware provides a standard basis for application program development on the AETMS platform, without having to be concerned with the complex network communication protocols. The following guidelines should be followed in establishing a standard middleware for the AETMS platform.

(1). The middleware should establish and maintain multiple parallel sessions on the platform network. The need for such multiple sessions arises when either a single client or server runs multiple applications or a single application communicates with other program modules across the network simultaneously. The multiple sessions are established by first setting up individual sessions (which are called TCP sockets in the TCP/IP world) and then linking them up into inter-process communications or IPCs.

(2). The middleware should replace the network specific application program interfaces (API). These interfaces are used to pass commands among the application programs during a multiple session in a transparent way. SQL adapted for multidatabase/program interaction is one of the ways in which this facility can be provided.

(3). The middleware should entirely manage all file transfers among the stations in the network. Some of the important utilities in file transfer systems include timed batch jobs, multiple simultaneous transfers, format conversion and error detection and correction.

(4). The middleware should entirely manage all message deliveries on the platform. This includes synchronous/asynchronous message passing, message queuing and distributed messaging. The middleware should facilitate an application to generate a message to another, forward it to a message queue and resume its own processing activity. The middleware should also facilitate message broadcast requirements. The middleware should incorporate queuing optimization models to control network traffic and manage the message routing process.

(5). The middleware should provide all these functional supports within a LAN, between LANs in a segment, and between segments. As a result, application program development and implementation should proceed without any concern for network communication management. The middleware serves as a black box, with well defined inputs and outputs for all these support requirements.

2.4 Platform Integration

The three layers constitute the foundations of the AETMS platform. However, an important question still remains: How are the three layers integrated such that the platform works as an unified whole. Platform integration involves three broad design issues: the fragmentation of the databases into partitions and replications, the allocation of the fragments among the stations in the network, and the design of a platform management protocol to co-ordinate the various tasks using the network middleware. Fragmentation and allocation define a part of the distributed application layer, and are based on the underlying network layer, user distribution and the application systems resident on the platform. The co-ordination mechanisms are the logical interfaces among the application layer, network layer and the users, and employ the middleware layer to regulate traffic within the network. We develop a design model of the co-ordination mechanisms in the following discussion.

Figure 3 shows a mapping of the data and network hierarchies presented earlier. The mapping is *one-to-many* between (global environment : local environments), (local environment : application systems) and (application system : application modules). Similarly, the mapping is *one-to-many* between (command network : segments) and (segment : LANs). The mapping of a LAN to its (server, clients) is *one-to-(one, many)*. The global environment covers the entire command network

and their mapping is *one-to-one*. The mapping (local environment : segment) is *many-to-many*, since a segment could contain several local environments and a local environment could span several segments. Similarly, the mapping (system : LAN) is *many-to-many*, since a LAN could contain several systems (at least part of them) and a system (at least parts of it) could be installed on several LANs. The same mapping applies to modules and (server, clients). If we consider the AETMS platform at the level of its basic *building blocks*, it is essentially made up of a collection of modules installed among a set of connected client/server stations throughout the command. As a result, the modules, stations and their links are *physical* entities, while the rest are *logical*. Hence, their respective mappings are also physical and logical, respectively. Using these mappings, we develop a strategy for fragmentation and allocation as follows.

Consider a logical segment. In designing a distributed database architecture on this segment, the following provisions must be made for the preservation of modularity and interoperability.

(a). The global schema should be maintained in the segment.

(b). The schema of each local environment should be maintained among its systems.

(c). The schema of each system should be maintained among its modules.

(d). The schema of a system should include logical and physical access paths to each of its modules and the structure of its database fragmentation. Similarly, the schema of a local environment should include the logical and physical access paths to each of its systems, and the global schema should include access path details to all the local environments within a segment and access hooks between segments.

(e). Database fragmentation involves vertical/horizontal partitioning of a database and replications. The following guidelines should be used in fragmentation: (i) Fragmentation should preferably not occur over segments. As far as possible, all the fragments of a database should be self-contained within a segment, (ii) Fragmentation should occur at the level of modules. A module can be partitioned/replicated among different LAN servers, and fragmentation within a LAN should be such that a client has a full or partial replication of the data held by its server, (iii) The distribution of the fragments should be consistent with the locality of reference among the user groups and the functional similarities among the modules of a system. Designing fragments as above would preserve modularity, trackability of fragments, efficiency in distributed query processing and integrity management and regulated network traffic.

While each of the above provisions can be made either centrally or distributively or in any combination, there are definite tradeoffs. Too much centralization could lead to considerable network traffic imbalances and inefficient use of network resources. Too much distribution could lead to excessive and avoidable workloads throughout the network, render the network chaotic, and hence,

result in inefficient use of resources. Between the lopsided and chaotic configurations, an optimal configuration for making these provisions exists. A fragmentation/allocation optimization model should be developed in order to arrive at these design decisions. However, for the sake of simplicity and clarity, we propose the following design architecture.

(1). The global schema is maintained by a non-user workstation such as a mini or a mainframe computer in the segment. This workstation is designated as the **AETMS Manager** of the segment.

(2). The schema of a local environment is maintained by a non-user workstation in one of its system LANs. This workstation is designated as the **Local Manager** of the environment.

(3). The schema of a system is maintained by all the servers of the LANs in which some of the modules of the system are stored. This is important for both immediate and total interoperability among the modules of the system. These servers are designated as the **System Managers** of the system. Due to the many-to-many relationship between systems and LANs, it is easy to envision a server acting as a system manager for several systems and a system having several managers.

The hierarchy of managers is shown in Figure 4, and could correspond to the actual organizational hierarchy within AETC. The managers are the *logical* interfaces between the application programs and the network middleware. They employ the middleware to access and transmit data from various locations to the location from which an application program requests data. We illustrate the co-ordination among the managers using a simple example as follows. To begin with, some notation is necessary. Let D_{ijkl} denote the lth module of the kth system in the jth local environment in the ith segment. Let M_{ijkl} denote the lth manager of the kth system in the jth local environment in the ith segment. Let N_{ij} denote the manager of the jth local environment in the ith segment. Let P_i denote the AETMS manager of the ith segment. Assume that an user at some client station X poses a query. This query will be handled as follows. Assume that processed data from modules D_{1111}, D_{1112}, D_{1121}, D_{1211} and D_{2111} should be obtained, assembled and refined to answer the query. Let the workstation X belong to a LAN that hosts D_{1111} entirely, and let its server be M_{1111}. Initially, X determines if the required data is locally available. It secures whatever data of D_{1111} is available, and passes on a request to M_{1111} for the missing data. Using the system schema, M_{1111} secures the remaining data of D_{1111} and determines the network access paths to servers hosting D_{1112} since it belongs to the same system. However, the data sets from D_{1121}, D_{1211} and D_{2111} can still not be determined, since they are not described in the system schema M_{1111} is holding. M_{1111} determines an access plan to obtain D_{1112} from its holding servers and passes on a request to its local environment manager for the remaining data. The access plan is arrived at using an

assessment of the existing network traffic conditions and the workloads of the potential servers of D_{1112} via the network middleware. A request-respond protocol is used for this purpose.

Let N_{11} be the local manager contacted by M_{1111}. Using its local schema, N_{11} determines that D_{1121} can be obtained from another system under its environment and chooses an appropriate access plan to arrange for the delivery of the required data from a holding server to M_{1111}. The request-respond protocol via the middleware is used for this purpose as well. Further, N_{11} also determines that D_{1211} and D_{2111} are not available within its domain, and refers them to its AETMS manager. Let P_1 be the AETMS manager contacted. Using the global schema, P_1 determines that D_{1211} can be obtained from another local environment within its domain and that D_{2111} has to be requested from outside. Accordingly, P_1 chooses an appropriate access plan for D_{1211} within its domain and contacts another AETMS manager (P_2, for example) who holds D_{2111}. P_1 arranges for the delivery of the required data from these sources to M_{1111}. The above sequence of co-ordinated transactions is summarized in Figure 5.

In the above scheme, the server of the LAN in which the query arises is chosen to assemble the data. However, alternate assembly patterns are possible, depending on the traffic intensity on the network and the server workload distributions. Further, decisions have to be made at servers M_{1111}, N_{11}, P_1 and P_2 to determine the optimal access plans. Hence, in designing a distributed operation such as this, efficient network routing optimization models should be available to support these decisions. This example intuitively develops the concepts behind the co-ordination mechanisms on the AETMS platform. We generalize and summarize these concepts in the communication path model shown in Figure 6.

The communication path model provides a basis for designing the co-ordination mechanisms on the AETMS platform. However, specific design optimization models for configuring these mechanisms should be developed. The optimization models should address the following design issues.

(1). How should the databases in each local environment be fragmented. If a module is partitioned/replicated, then how many partitions/replications should there be, and where should they be stored and managed.

(2). What is the structure of the various schemata used in the distributed architecture. Where should they be stored and how are they used in communication.

(3). Who are the designated managers. How does this designation correspond to the organizational structure. What is the overall schemata management plan (centralized/distributed/combined).

(4). How is the communication path determined from the schemata management plan. How and where the data for any query is assembled. Are the communication paths and assembly plan determined on the fly or fixed in advance.

(5). What is the likely effect of each design on performance measures such as query response times, workload distribution, network traffic intensity, transaction processing and integrity/security management performance.

It is usually difficult to address all these issues in a single comprehensive model. We suggest partitioning the issues into functionally related sets and address them using specific models. For example, the database fragmentation can be initially determined using a *fragmentation model*. The allocation of fragments and modules to network LANs/devices can then be determined using an *allocation model*. The fragmentation and allocation lead to the schemata management plan, which is determined by a *schemata design model*. The schemata structure and the database/module distribution lead to the design of the network management hierarchy, communication paths and data assembly procedures. This can be carried out using a *network management model*. All these models identify the design variables in a decision making framework and determine a design configuration that is optimal with respect to the system performance measures. The optimization is carried out using analytical modelling and computer simulations of real-time platform operations.

The proposed architecture of the integrated AETMS platform is shown in Figure 7. This architecture shows the components of the platform in terms of the three layers, their attributes and functionalities, and the integration of the components using the platform managers. The specifications of the proposed architecture and the underlying optimization models need to be developed and established as AETC standards for all future system development on the platform. Finally, one of the important issues in commandwide standardization of the platform is *benchmarking*. This is discussed as follows.

Benchmarking is a standardized methodology for measuring the performance of hardware and software systems in a controlled environment. Benchmarking is the basis for T&E, and should be applied at every stage of system development: requirements analysis, capacity planning, specification of all system requirements in RFP, contractor/vendor evaluation, system design, T&E procedures, system implementation and maintenance. While many public and proprietary benchmarks exist, a benchmark appropriate for the AETMS platform should be developed. The benchmarks should provide tools for assessing the performance of workstations, network communication systems and distributed database management systems. Workstation performance is usually assessed in terms of computing speed, main/cache memory, disk space and user interfaces. Network performance is assessed in terms of system connectivity, bandwidth/data rates and the performance of various management protocols. The distributed database systems are assessed in terms of input/output architecture, data paths, query response times, throughput effectiveness of concurrency control, import/export abilities, backup/restore abilities, system scalability, integrity/security architecture and overall performance under single and multiuser test conditions. In developing a set of benchmarks for the AETMS platform, the following should be taken into account. A benchmark should be

relevant. It should measure accurately the peak performance and price performance of a system under realistic workloads. It should be simple to use and scalable to a wide spectrum of computing systems. It should be portable and should not be biased in favor or against any vendor.

3 Conclusion

Aircrew training is an important function in the Air Force. Training is a complex and highly diversified large-scale operation requiring a considerable committment of resources. Integrated training management systems are intended to provide efficient information and decision support in various training related activities. AETMS is a concept derived from this doctrine and is aimed at supporting all training requirements at AETC from a central integrated information management platform.

The goals of this study are as follows: study the current operations at AETC and take the AETMS vision one step further. To this end, this study yielded the following results: a design model of the AETMS platform, an architecture of the suite-of-systems on the platform, a program plan for the short-run developmental tasks and a focus for the long-run activities. The immediate steps in advancing towards the AETMS goal are: (i) develop and establish commandwide AETMS platform standards starting from the proposed model, (ii) make the necessary provisions suggested in the comprehensive report submitted to AL/HRA and AETC for infrastructure development, and (iii) commence the short-run project with an eye towards the long term goal.

AETC WINGS

OTHER AF COMMANDS

NAVY

OTHER DoD AGENCIES

AUTODIN

DDN

AETC/HQ Segment

ROUTER

ROUTER

BACKBONE NETWORK

ROUTER

ROUTER

MODULAR CONCENTRATOR

ACTIVE FOIRL HUB

ACTIVE FOIRL HUB

MOD MOD MOD MOD MOD MOD

MOD MOD MOD MOD MOD MOD

TP

coax

fiber

DEVICES (LANs)

PASSIVE HUB

PASSIVE HUB

coax

fiber

PC

PC

PC

PC

PC

DEVICES (LAN)

DEVICES (LAN)

(LAN)

(LAN)

(LAN)
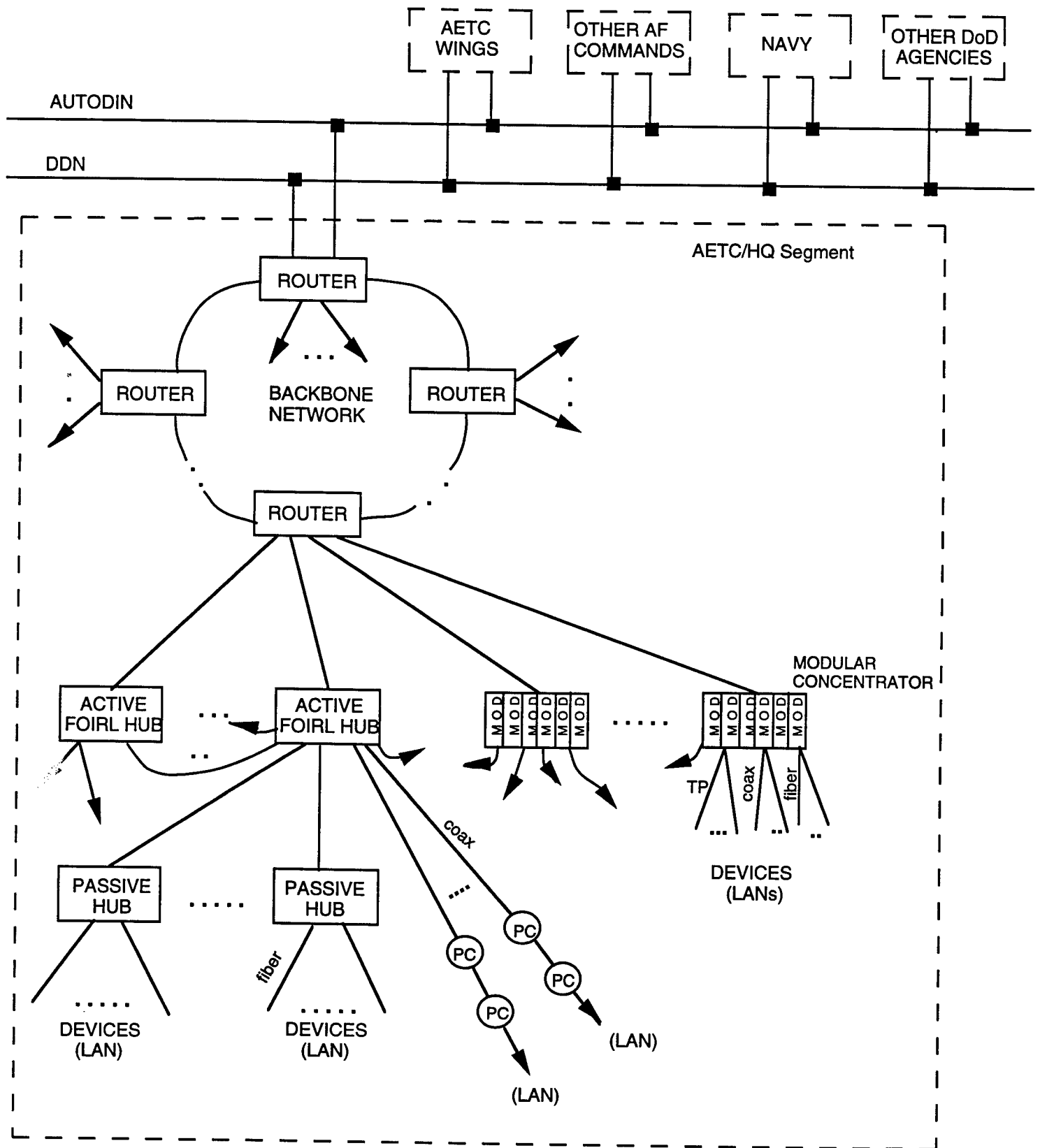
FIGURE 1 Physical Host Architecture
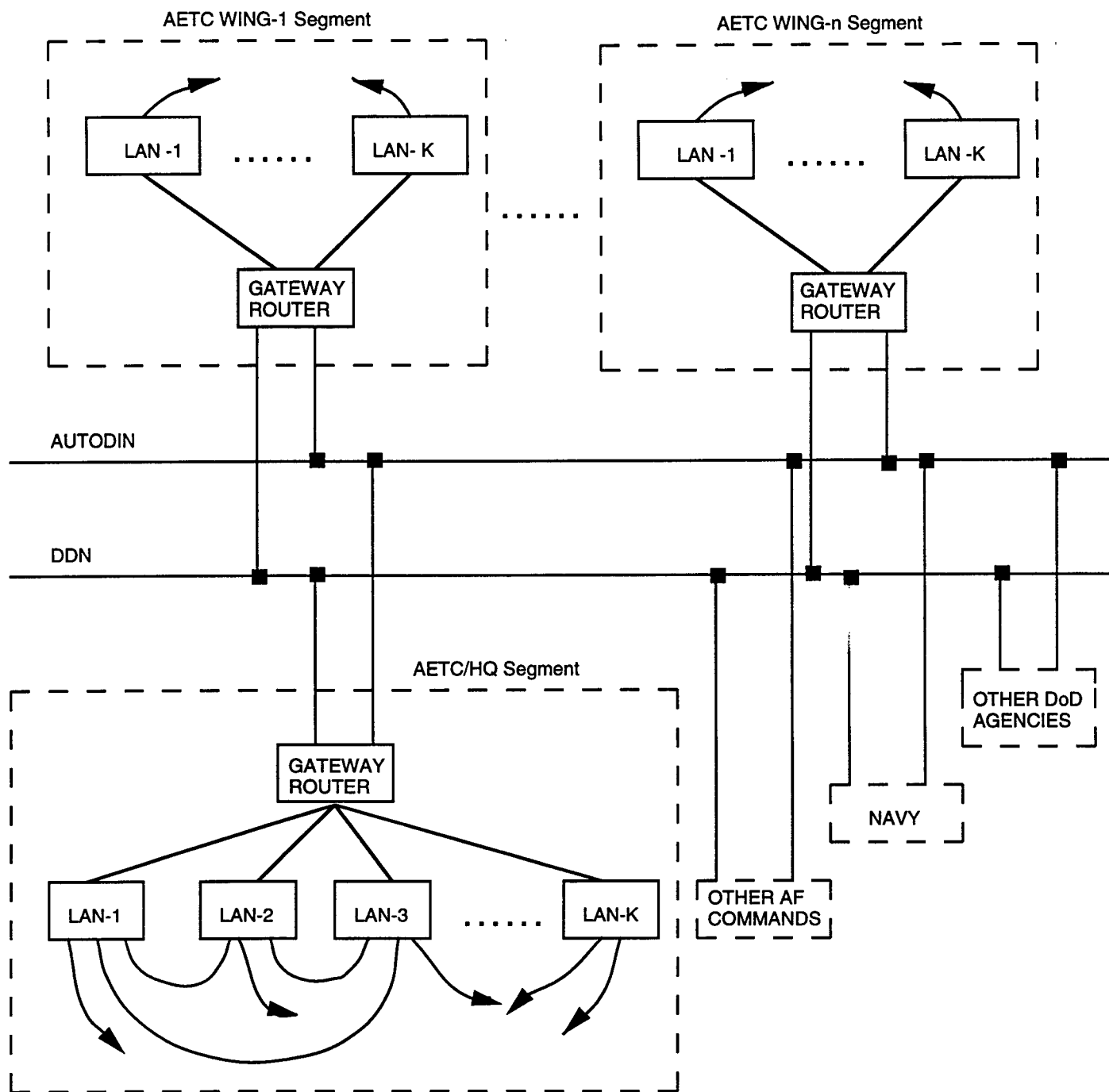
FIGURE 2 Logical Host Architecture

Data hierarchy

Network hierarchy

Global Environment (1: n) — 1:1 → Commandwide Network (1: n)

Local Environment — Local Environment — m:n → Segment (1:n) — Segment

Local Environment (1:n) — System

System (1:n) — System — m:n → LAN (1:(1,n)) — LAN

Module — Module — m:n → Server — Clients

All mappings are logical

Physical mapping
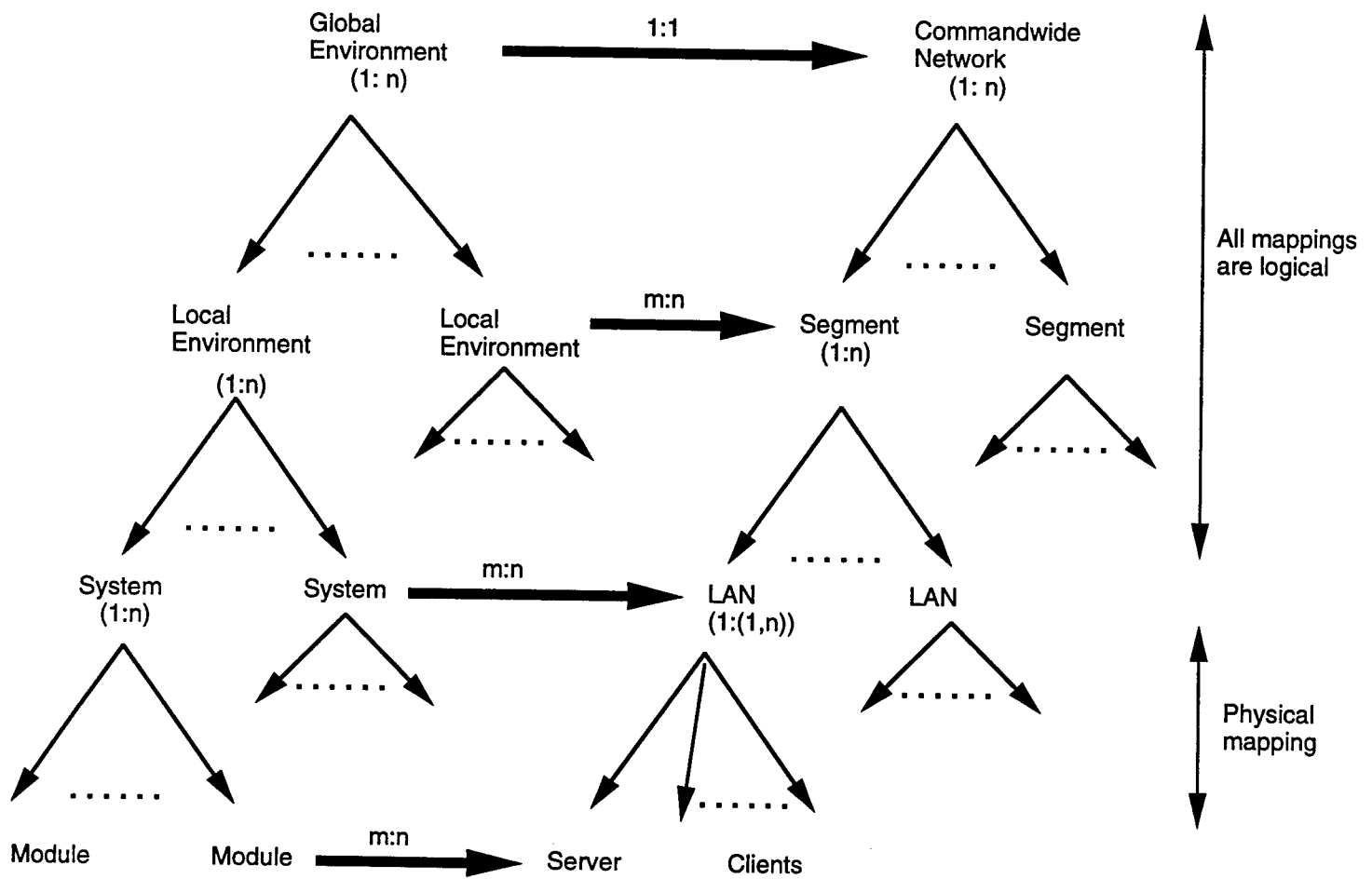
FIGURE 3 Mapping of hierarchies

35 - 16
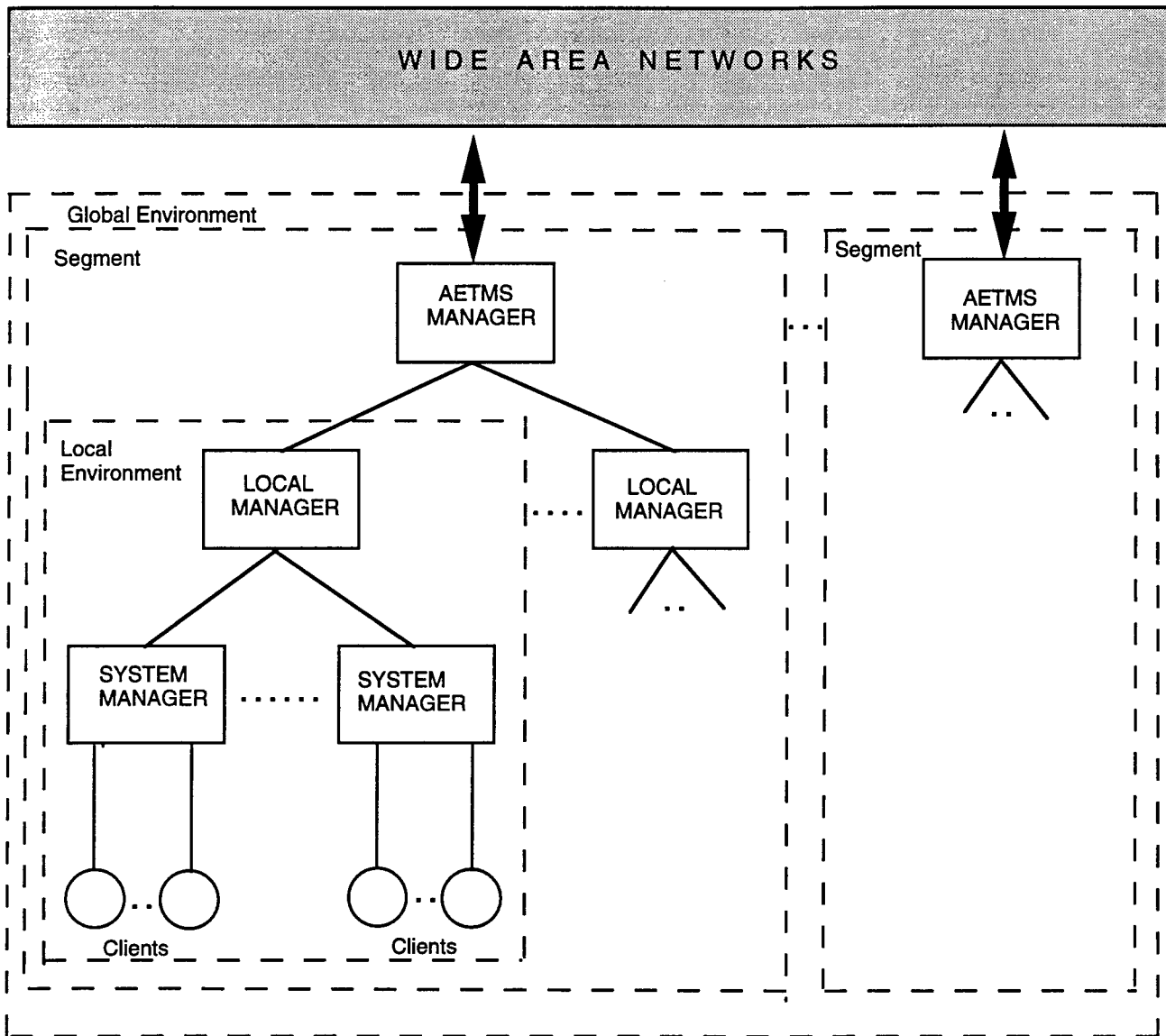
FIGURE 4 Hierarchy of Managers
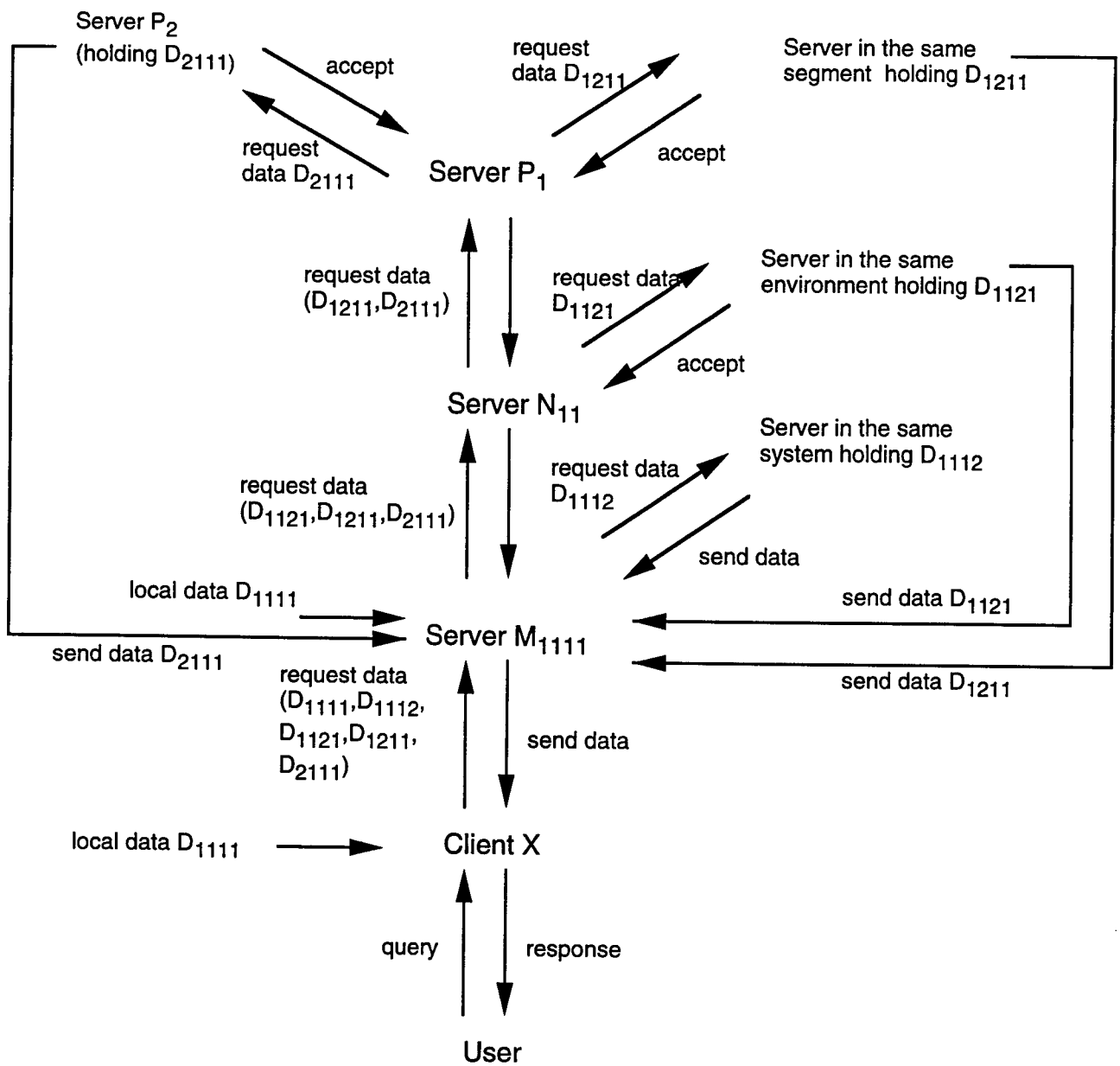
FIGURE 5 Coordination among Distributed Stations

FIGURE 6 Communication Path on the AETMS platform
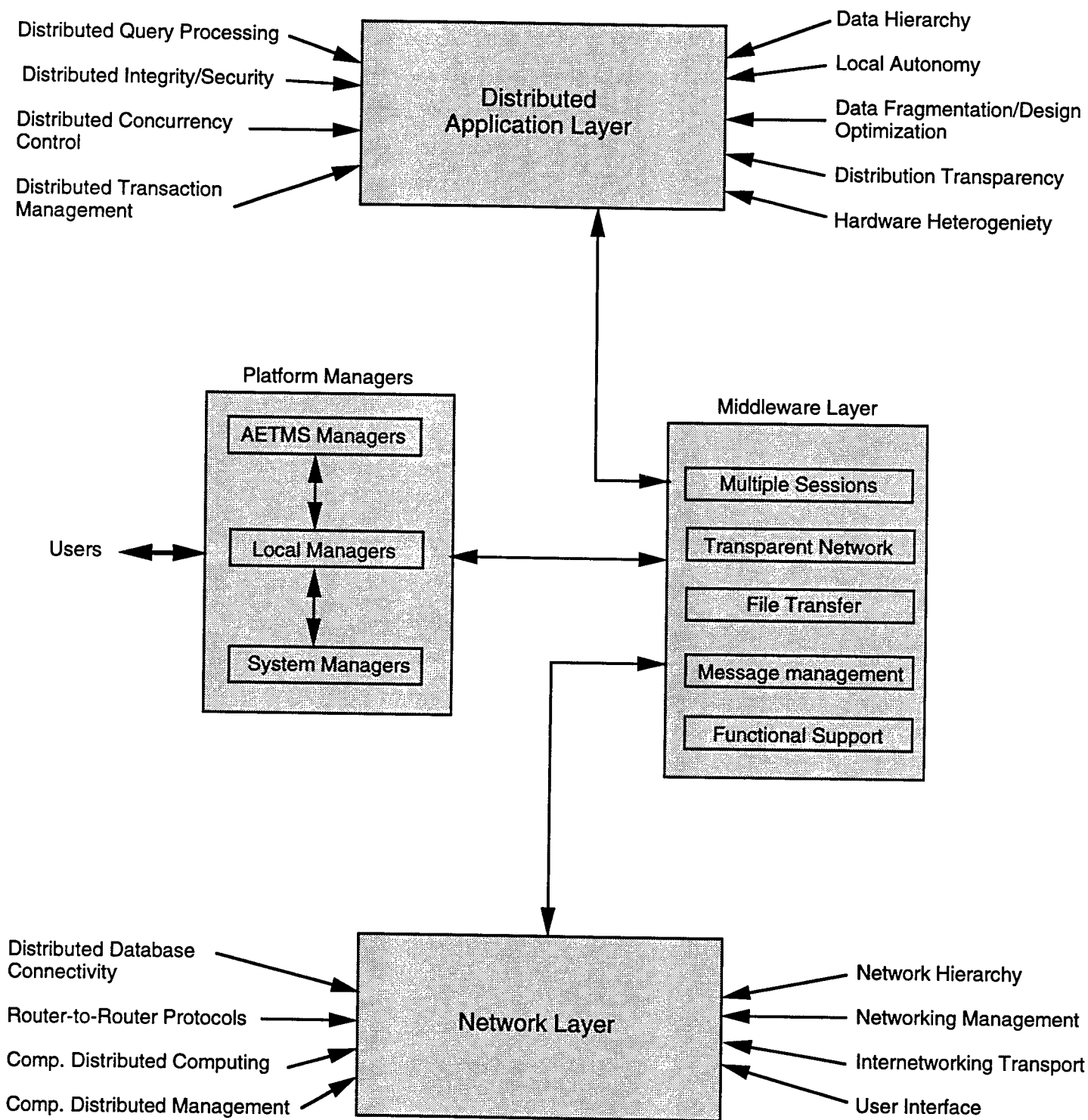
FIGURE 7 Integrated AETMS Platform Architecture

Dr. Gary E. Riccio
Report Unavailable

**SEQUENTIAL ESTIMATION OF PARAMETERS OF
TRUNCATION PARAMETER FAMILIES**

AND

HALF-LIFE STUDIES IN RANCH HAND VETERANS

Kandasamy Selvavel
Assistant Professor
Department of Mathematics
and
Computer Science

Claflin College
700 College Avenue
Orangeburg, SC 29115

TABLE OF CONTENTS

SEQUENTIAL ESTIMATION OF PARAMETERS OF TRUNCATION PARAMETER FAMILIES

AND

HALF-LIFE STUDIES IN RANCH HAND VETERANS

Kandasamy Selvavel
Assistant Professor
Department of Mathematics and Computer Science
Claflin College

Abstract

In the first part of this study, we considered a one truncation parameter family of distributions with pdf of the form $f(x;\theta)=q(\theta)h(x)$, $a<x<\theta$. We studied the sequential estimation of a function of the unknown parameter under type II censoring. The stopping rule is derived in a manner to minimize the total of the risk plus the cost of sampling. We showed that the sequential estimator is asymptotically unbiased. Also, we proved that when $b>0$, the stopping rule terminates finitely with probability one. Moreover, it is also shown that the stopping rule is asymptotically efficient and risk efficient as b tends to zero.

In the second part of this study, we found a new half-life of dioxin concentration in Ranch Hand veterans. First, we estimated the regression effect on decay rate, and then calculated the corrected half-life using the regression effect. Using 1982 and 1987 dioxin concentrations the corrected half-life was found to be 9.0 years, and using 1987 and 1992 dioxin concentrations it was found to be 7.3 years. Using all three points, that is 1982, 1989 and 1992 dioxin concentrations, the half-life was estimated as 7.77 years.

SEQUENTIAL ESTIMATION OF TRUNCATION PARAMETER FAMILIES

Kandasamy Selvavel

CHAPTER I

1.1 Introduction

The truncated(nonregular) distributions are those distributions for which one or both of the extremities of the range are functions of the parameters. Some of the well-known distributions are of this type. This is the case, for example, for the uniform distribution with probability density function (pdf) $1/\theta$ on $[0, \theta]$, the exponential pdf $\exp\{-(x-\mu)\}$, $x>\mu$, and the Pareto pdf $\alpha\sigma^{\alpha}/x^{\alpha+1}$, $x>\sigma$.

Several authors have studied the merits of the truncation parameter families. These families were first studied by Davis (1951), and later by many authors including Guenther (1978), Bar-Lev and Boukai (1985), Ferentinos (1987), Krishnamoorthy, Rohatgi and Blass (1989), Rohatgi (1989) and Selvavel (1989).

In the case of truncation parameter family of distributions, derived functions, such as the survival and hazard functions, depend on truncation parameters. As a result of this, there is interest in estimating parameters of these families.

Davis (1951) and Guenther (1978) considered pdfs of the form $f(x;\theta)=q(\theta)h(x)$, $a<x<\theta$ and $f(x;\theta) = q(\theta)h(x)$, $\theta<x<b$, and obtained uniform minimum variance unbiased (UMVU) estimators of the unknown parameters. Krishnamoorthy, Rohatgi and Blass (1989) considered unbiased estimation in type II censored samples from these families. Selvavel (1989) extended this result to the two sample-case.

The problem of estimation of parameters in truncation parameter family of distributions has been well studied when complete samples are available. But, only a few results are known for censored and sequential cases. In our study, we consider the sequential estimation of a function of truncation parameter of a one-truncation parameter family of distributions under type II censoring. In this sequential estimation, we use a stopping rule to choose the sample size in a manner that minimizes the risk plus the cost of sampling. Wald (1947) was the first to suggest the use of stopping rule in the sample size selection. Since then several authors, notably Anscombe (1953), Chow and Robbins (1965) and Woodroofe (1982) have done extensive work on sequential estimation problems. In this study, we look at the theoretical merits of the stopping rule and the sequential estimate of a function of truncation parameter family of distributions under type II censoring.

1.2 Notations and Assumptions

Let us suppose that the observations x_1, x_2, x_3, ..., x_n follow a one-truncation parameter family of distributions with probability density function (pdf)

$$f(x;\theta)=\begin{cases} q(\theta)h)x), & a^* \leq x < \theta \leq b^* \\ 0, & \text{otherwise,} \end{cases} \qquad (2\text{-}1)$$

where $-\alpha < a^* < b^* < \alpha$ are known constants, $q(\theta)$ is everywhere differentiable, and $h(x)$ is positive and absolutely continuous. Our objective is to minimize total cost, $T=a(\text{risk})+b\log(r)$, where r is the number of available observations. Here,

$\text{risk} = E[\hat{g}(\theta)-g(\theta)]^2$ and $\hat{g}(\theta)$ is the maximum likelihood estimate of $g(\theta)$. In this case,

$$\hat{g}(\theta)=\hat{g}(x_{r:n}).$$

We assume that

(I) a^* and b^* are known real numbers.

(II) a and b are real numbers such that $a=0(b)^{1-s}$, where $0<s<1$.

(III $q(\theta)$ is differentiable and bounded away from zero.

(IV) $r = [np] \geq 4$

(V) $g(\theta)$ is differentiable.

1.3 Main Results

Let x_1, x_2, x_3, ..., x_n be a random sample from a population having probability density function given by (2-1) with $x_{i:n}$ denoting the ith order statistic. Suppose that the sample is type

II censored and that one observes only $x_{1:n}$, $x_{2:n}$, $x_{3:n}$, ..., $x_{r:n}$, $1 \leq r \leq n$. Our objective is to estimate $g(\theta)$, a function of the truncation parameter θ in such a way that $T=a.E[q(x_{r:n})-g(\theta)]^2+$ $b\log(r)$, is minimized.

We need the following lemma in order to proceed with the development of the theory.

Lemma 1: Let $g(\theta)$ be a continuous function in θ. Then there exists a continuous function $H(\theta)$ such that $g(\theta)=H(g(\theta))$.

Proof: Note that
$$q(\theta)=[\int_{a}^{\theta} h(u)\,du]^{-1}.$$

By differentiating both sides with respect to θ, we get

$$q'(\theta)=-q^2(\theta)h(\theta)<0.$$

Therefore, $q(\theta)$ is one-to-one and invertible.
We note that
$$g(\theta)=gq^{-1}(q(\theta))=H(q(\theta)),$$
where $H=gq^{-1}$, and it is continuous. It can be easily proved that
$$E[g(x_{r:n})-q(\theta)]^2=q^2(\theta)\{n(n-1)/[(r-1)(r-2)]-2n/(r-1)+1\}.$$
Now the focus changes to minimizing
$$T=aE(H(q(x_{\gamma:n}))-H(q(\theta)))]^2+b\log(r)$$
$$\simeq a[H'(q(\theta))]^2E[q(x_{r:n})-q(\theta)]^2+b\log(r)$$
$$=a[H'(q(\theta))]^2q^2(\theta)\{n(n-1)/[(r-1)(r-2)]-2n/(r-1)+1\},$$
where a and b are two given real numbers such that $a=0(b^{1-s})$, for some s such that $0<s<1$.
By differentiating T with respect to n, and ignoring terms of order $O(1/n)$ and $O(1/n^2)$, we get
$$n=2/p+\{[a(1-p)]/bp^3\}[H'(q(\theta))]^2$$
So, the stopping rule is, stop the sampling when

$$N_0 = \inf\{n : n \geq 2/p + \{[a(1-p)]/bp^3\}[H'(q(\theta))]^2\} \tag{3-1}$$

If θ is unknown, use the rule

$$N = \inf\{n : n \geq 2/p + \{[a(1-p)]/bp^3\}[H'(q(X_{r:n}))]^2\} \tag{3-2}$$

Lemma 2: $N_0 \to \alpha$ as $b \to 0$.

Proof: Since $a = 0(b^{1-s})$, we obtain that $a/b = 0(b^{-s})$ and $a \to \alpha$ as $b \to 0$. So, $N_0 \to \alpha$ as $b \to 0$.

Lemma 3: $N \to \alpha$ as $b \to 0$.

Proof: Similar to the proof of Lemma 1.

Lemma 4: $X_{r:n} \to \xi_p$ a.s. as $n \to \alpha$, where $r = [np]$ and ξ_p is the pth quantile of $f(x:\theta)$.

Proof: The result follows from Bahadur's Theorem (1966) and Glivenko-Cantelli Lemma.

Lemma 5: $X_{r:N} \to \xi_p$ a.s as $N \to \alpha$, where $r = [Np]$.

Proof: Let $Z_{n(b)} = X_{[pn(b)]:n(b)}$. By Bahadur's Theorem $Z_{n(b)} \to \xi_p$ a.s. as $b \to 0$, and by Csorgo's Theorem (1963) $Z_{N(b)} \to \xi_p$ a.s. as $b \to 0$. That is, $X_{pN(b):N(b)} \to s_p$ a.s. as $b \to 0$. Hence the result.

Lemma 6: $q(X_{pN(b)]:N(b)}) \to q(\xi_p) = (1/p)q(\theta)$

and

$$q^2(X_{pN(b):N(b)}) \to q^2(\xi_p) = (1/p^2)q^2(\theta) \text{ a.s. as } b \to 0.$$

Proof: Follows from the Continuous Capping Theorem.

Lemma 7: $E[q(X_{bN(b)]:N(b)})] \to \xi_p$ as $b \to 0$.

Proof: Note that $q(X_{[p.N(b)]:N(b)}) \leq$ maximum of $q(y)$ for $y > a^*$, and hence by the Dominated Convergence Theorem the result follows.

Lemma 8. $\{([Np]-1])/N\}q(X_{[Np]:N}) \to q(\theta)$ a.s. as $b \to 0$.

Proof: Follows from Lemma 5 and Slutsky's Theorem.

Lemma 9: $E\left\{\{([Np]-1])/N\}q(X_{[Np]:N)}\right\} \to q(\theta)$ as $b \to 0$.

Proof: Note that

$$([Np]-1])/N \leq 1$$

and

$$q(X_{[Np]:N}) \leq \text{maximum of } q(y) \text{ for } y > a^*.$$

Hence, $\{([Np]-1])/N\}q(X_{[Np]:N}) \leq$ maximum of $q(y)$ for $y > a^*$, and then by the Dominated Convergence Theorem the result follows.

Lemma 10: $N/N_0 \to \{[g'(\xi p)]/g'(\theta)\}^2$ a.s. as $b \to 0$.

Proof: The result follows from (3-1), (3-2) and Lemma 5.

Lemma 11: $E(N/N_0) \to \{[g'(\xi p)]/g'(\theta)\}^2$ as $b \to 0$.

Proof: Follows from the Dominated Convergence Theorem and Lemma 10.

Lemma 12: Let

$$T^{\wedge} = aE[g(X_{r:N} - g(\theta)]^2 + bE(\log(r))$$

and

$$T = aE[g(X_{r_0:N_0}) - g(\theta)]^2 + b\log(r_0), \text{ where } r = [Np] \text{ and }$$

$r_0 = [Np]$, then $(T^{\wedge}/T) \to 1$ as $b \to 0$.

Proof: Using the definition of $T^{\wedge}$ and T, it can be easily shown that $\{(T^{\wedge}-T)/T\} \to 0$ as $b \to 0$. That is, $(T^{\wedge}/T) \to 1$ as $b \to 0$. Hence the result

Lemma 13: $E(N) = \lambda + 2/p + \{[a(1-p)]/bp^3\}[g'(\theta)]^2$

$$+ (a/b)O(\sigma_{r:N}) + (a/b)O\{(\sigma_{r-1:N-1})^2 + (\sigma_{r:N})^2\}^{1/2},$$

where $\sigma_{r:N}^2 = E[X_{r:N} - \theta]^2$.

Proof: From (3-2), we have

$$N = 2/p + \{[a(1-p)]/bp^3\}[g'(\theta)]^2$$

$$+ \lambda\{1 + \{[a(1-p)]/bp^3\}\{[g'(X_{r-1:N-1})]^2 - g'(X_{r-1:N-1})]^2\} \qquad (3-3)$$

The result follows by taking expected values on both sides of (3-3), and then applying the Mean Value Theorem and the Cauchy-Schwartz Inequality to the right hand side of the equation.

Lemma 14: $N_0 = 2/p + \{[a(1-p)]/bp^3\}[g'(\theta)]^2 + \lambda^{\sim}$, for some $0 < \lambda^{\sim} < 1$.

Proof: Follows directly from (3-1).

Lemma 15: $E(N) = N_0 + \lambda^* + (a/b)O\{(\sigma_{r-1:N-1})^2 + (\sigma_{r:N})^2\}^{1/2}$, for some $-1 < \lambda^* < 1$.

Proof: Follows from Lemma 13 and Lemma 14.

Lemma 16: The stopping rule given by (3-2) terminates finitely with probability one when b>0.

Proof: Note that when b>0, we have a>0. Also $0<p<1$ and $[g'(X_{r:n})]^2$ is finite. Therefore, the right hand side of (3-2) is finite. Hence the result.

Lemma 17: $E[g(X_{r:N})]=g(\theta)+O(\sigma^*_{r:N})$, where $(\sigma^*_{r:N})^2$ is the mean square error of $X_{r:N}$.

Proof: By Mean Value Theorem

$$g(X_{r:N})-g(\theta)=(X_{r:N}-\theta)g(s), \text{ for some S such that } X_{r:N}\leq S\leq\theta.$$

Taking expected value on both sides, we get

$$E\{g(X_{r:N})-g(\theta)\}=E\{(X_{r:N})-g(\theta)\}. \tag{3-4}$$

Using the Cauchy-Schwarz inequality, we have

$$E(|g(X_{r:N})-g(\theta)|)\leq M[E(X_{r:N}-\theta)^2]^{1/2}, \text{ where } M=\sup_{0\leq x\leq\theta} g'(X)$$

$$= M(\sigma^*_{r:N}). \tag{3-5}$$

The result follows by combining (3-4) with (3-5).

Remark: As $b\to0$, $N\to\alpha$. Therefore, $g(X_{r:N})$ is an asymptotically unbiased estimator of $g(\theta)$.

1.4 Discussion

In this study, we considered a one-truncation parameter family of distributions with pdf of the form $f(x;\theta)=q(\theta)h(x)$, $a< x<\theta$. This family of distributions has useful applications in biometry, engineering and natural sciences. We studied the sequential estimation of a function of the unknown truncation parameter under type II censoring. We assumed that in a sample of n observations, only a proportion p of the observations are available for the study. The stopping rule is derived in a manner to minimize the total of the risk plus the cost of sampling. We showed that the sequential estimator of a function

of the parameter is asymptotically unbiased. Also, we proved that

when b>0, the stopping rule terminates finitely with probability

one. Moreover, it is shown that the stopping rule is

asymptotically efficient and risk-efficient as b tends to zero.

These results may be generalized to two-truncation parameters

families.

CHAPTER 2

REGRESSION TOWARDS THE MEAN IN HALF-LIFE STUDIES

2.1 Introduction

The estimation of the half-life of dioxin in Vietnam veterans of Operation Ranch Hand was first studied by Pirkle et al. (1989). Thirty-six pairs of observations with at least 10 parts per trillion (10 ppt) concentration of dioxin were selected for their study. They used the standard half-life equation derived from first-order kinetics to estimate the half-life. Using a background correction of 4 ppt, they obtained median half-life as 7.1 years with a 95% confidence interval of 5.8 - 9.6 years. Michalek et al. (1992) studied the relationship between TCDD half-life and percentage of body fat and changes in percentage of body fat of the same group of 36 Air Force veterans that were considered by Pirkle et al. (1989). They used a repeated-measures analysis of a covariance model $y_{ij}=\beta_0+\tau_i+\beta_1 t_{ij}+\epsilon_{ij}$, where y_{ij} is the natural logarithm of the jth TCDD background corrected measurement of the ith individual, t_{ij} is the number of years between the jth and first TCDD measurements, $-\beta_1$ is the common decay rate, $(\beta_0+\tau_i)$ is the intercept for the ith individual, and ϵ_{ij} is the residual error term for measurement y_{ij}. The half-life corresponding to this model was estimated as 7.68 years with a

95% confidence interval of 6.2-10.1 years. They also found that the half-life has only a marginal relationship with percentage of body fat. Moreover, they showed that there was no significant relationship between half-life and relative change in percentage of body fat. Recently, Wolfe et al. (1994) estimated the half-life as 11.6 years using 337 Ranch Hand veterans. They also found that in the regression model, age, percentage of body fat (PBF), and the relative change in PBF over a 5-year period were all significant predictors of dioxin half-life.

In our study we estimated the corrected half-life by finding the regression effect on decay rate. The concept of regression towards the mean was first introduced by Galton almost a century ago. However, very few standard results are known in the literature, and these results have not yet been introduced in the text books. In this study, we first estimated the regression effect using two time points, and it was found as the products of k, σ and $(1-\rho)$, where $k=\phi(C)/[1-\Phi(C)]$. Here, C is the standardized cut point and ϕ, Φ are the standard normal density and cumulative distribution functions respectively. In the second approach we estimated the regression effect using three time points. Then the corresponding half-lives were calculated using these regression effects. In both approaches, we assumed that the data follows a multivariate normal distribution.

2.2 Methods

(a) Regression Effect on Decay Rate

Let X_1 and X_2 be the random variables corresponding to the logarithms of 1987 and 1992 dioxin concentration in the Ranch Hand veterans respectively. We assume that $X = (X_1, X_2)'$ is normally distributed with mean vector $\mu = (\mu, \mu-\tau)'$, $\tau > 0$, and covariance

matrix $\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ (2-1)

Let $D = X_1 - X_2$, and ΔT be the time from 1987 dioxin measurement until 1992 dioxin measurement in unit of years. Furthermore, let c_0 be the left truncation point of logarithm of 1987 dioxin concentration.

Theorem 2.1: Let D, c_0 and ΔT be defined as above. Then the regression effect on estimating decay rate of dioxin concentration for the above model is given by $\{k_0 \sigma (1-\rho)\}/\Delta T$, where $k_0 = \phi(C_0)/[1-\Phi(C_0)]$ with $C_0 = (c_0 - \mu)/\sigma$, where ϕ and Φ are the standard normal density and cumulative distribution functions respectively.

Proof: Note that

$$E(D \mid X_1 > c_0) = E(X_1 - X_2 \mid X_1 > c_0)$$
$$= \mu + \sigma \phi(C_0)/[1-\Phi(C_0)] - \{\mu - \tau + \sigma\rho\phi(C_0)/[1-\Phi(C_0)]\}$$
$$= \tau + \sigma(1-\rho)\phi(C_0)/[1-\Phi(C_0)],$$
$$= \tau + \sigma k_0(1-\rho), \qquad (2-2)$$

where $C_0 = (c_0 - \mu)/\sigma$.

The uncorrected decay rate is

$$\lambda_u = E(D/\Delta T \mid X_1 > c_0).$$

and the true decay rate is $\lambda_t = \tau/\Delta T$.

Therefore, from (2-3) the true decay rate can be written as

$$\lambda_t = \lambda_u - \sigma k_0(1-\sigma)/\Delta T.$$

Hence, the regression effect on estimating decay rate is $\sigma k_0(1-\rho)/\Delta T$.

Remark 2.1: Note that the regression effect approaches to zero when ρ tends to one.

Next, we find the decay rate using 1982 and 1987 dioxin concentrations. Let X_1 and X_2 be the logarithm of 1982 and 1987 dioxin concentrations respectively. Furthermore, assume $X = (X_1, X_2)'$ is normally distributed with mean vector $\mu = (\mu + \tau, \mu)'$, $\tau > 0$, and covariance matrix Σ given by (2-1). Also, let $D = X_1 - X_2$ and c_0 be the left truncation point of 1987 dioxin concentration. We define ΔT as the time from 1982 dioxin measurement until 1987 dioxin measurement in years.

Theorem 2.2: Let D, c_0 and ΔT be defined as above. Then the regression effect on estimating decay rate of dioxin concentration for the above model is given by $\{k_0\sigma(1-\rho)\}/\Delta T$, where $k_0 = \phi(C_0)/[1-\Phi(C_0)]$ with $C_0 = (c_0 - \mu)/\sigma$, where ϕ and Φ are standard normal density and cumulative distribution functions respectively.

Proof: Similar to that of Theorem 2.1.

We now truncate on both variables X and Y, and estimate the regression effect on decay rate. Let Z be a normally distributed random vector with mean μ and covariance matrix Σ, where $Z = (X,Y)'$, $\mu = (\mu, \mu - \tau)'$, $\tau > 0$, and Σ given by (2-1). It can be easily shown that

$$E(X-Y \mid X > c_1, Y > c_2)$$
$$= \tau + \{\sigma(1-\rho)/P\}\{\phi(C_1) - \Phi(C_2) + \Phi(C_2)\Phi[(C_1-\rho C_2)/\sqrt{1-\rho^2}] - \Phi(C_1)\Phi[(C_2-\rho C_1)/\sqrt{1-\rho^2}]\},$$

where $P = P[X > c_1, Y > c_2]$, $C_1 = (c_1 - \mu)/\sigma$ and $C_2 = (c_2 - \mu + \tau)/\sigma$.

Remark 2.3: If $c_2 = -\infty$, we get

$$E(X-Y \mid X > c_1) = \tau + \sigma(1-\rho)\phi(C_1)/\{1-\Phi(C_1)\}$$
$$= \tau + k\sigma(1-\rho),$$

where $k = \phi C_1)/\{1-\Phi(C_1)\}$.

Note that this result coincides with that of Theorem 2.1.

Remark 2.4: If $c_1 = c_2 = c$, we get

$$E(X-Y \mid X > \mu + c\sigma, Y > \mu - \tau + c\sigma) = \tau.$$

That is,

$$E(X-Y|X>k,Y>k-\tau)=\tau,$$

where $k=\mu+c\sigma$.

Therefore, there is no regression effect when the two cut points differ by τ.

(b) Regression to the Mean on Estimating Slope in Ordinary Repeated Measures Linear Regression

In this section we consider a simple repeated measures linear model

$$y_{ij}=\beta_0+\beta_1 t_{ij}+\epsilon_{ij}, \quad 1\le i\le n, \quad 1\le j\le 3,$$

where y_{ij} is the natural logarithm of the jth dioxin measurement of the ith individual, t_{ij} is the number of years between the jth and first dioxin measurements, and β_1 is the slope of the regression line and ϵ_{ij} is assumed to be $N(0,\sigma^2)$ for all i and j. Then the weighted least squares estimate of β_1 can be written as

$$\hat{\beta_1}=(nS_{ty}-S_yS_t)/(nS_t2-S_t^2),$$

where

$$S_t = \sum_{i=1}^{n}\sum_{j=1}^{3} t_{ij}, \quad S_t = \sum_{i=1}^{n}\sum_{j=1}^{3} y_{ij}, \quad S_t^2 = \sum_{i=1}^{n}\sum_{j=1}^{3} t^2_{ij}, \quad \text{and } S_{ty} = \sum_{i=1}^{n}\sum_{j=1}^{3} t_{ij}\, y_{ij}.$$

Equivalently, $\hat{\beta}_1$ can be written as a weighted average of the y_{ij}. That is,

$$\hat{\beta_1} = \sum_{i=1}^{n}\sum_{j=1}^{3} w_{ij}\, y_{ij},$$

where $w_{ij} = (t_{ij} - t^-)/\sum_{i=1}^{n}\sum_{j=1}^{3} (t_{ij} - t^-)^2$.

Note that the sum of the weights is zero.

As a particular case suppose that

$$t_{i1} = t_i - \Delta, \quad t_{i2} = t_i, \quad t_{i3} = t_i + \Delta.$$

Then

$$\sum_{i=1}^{n}\sum_{j=1}^{3} (t_{ij}-t^-)^2=3(n-1)S_t2+2n\Delta^2, \quad W_{i1}=(t_i-t^--\Delta)/[3(n-1)S_t2+2n\Delta^2],$$

$$W_{i2}=(t_i-t^-)/[3(n-1)S_t2+2n\Delta^2], \quad w_{i3}=(t_i-t^-+\Delta)/[3(n-1)S_t2+2n\Delta^2].$$

Now we assume that Y is normally distributed with mean vector μ and covariance matrix Σ, where $Y = (Y_1, Y_2, Y_3)'$, $\mu = (\mu+\tau, \mu, \mu-\tau)'$ and

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

Then

$$E(\hat{\beta}_1 | Y_2 > c) = \sum_{i=1}^{n} \sum_{j=1}^{3} w_{ij} E(Y_{ij} | Y_2 > c)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{3} \{ w_{i1} E(Y_{ij} | Y_2 > c) + w_{i2} E(Y_{i2} | Y_2 > c) + w_{i3} E(Y_{i3} | Y_2 > c) \}$$

We need the following results in order to proceed further.

$E(Y_{i1} | Y_2 > c) = \mu + \tau + \rho\sigma k$, $E(Y_{i2} | Y_2 > c) = \mu + \rho\sigma k$ and $E(Y_{i3} | Y_2 > c) = \mu - \tau + \rho\sigma k$, where $k = \phi(C)/\{1 - \Phi(C)\}$, $C = (c-\mu)/\sigma$.

Using the above results and some simplification, we get

$$E(\hat{\beta}_1 | Y_2 > c) = (-\tau/\Delta) \{ 1/[1 + (3/2)((n-1)/n)(S_t2/\Delta^2)] \}.$$

Therefore, the proportion of the regression effect on the slope is $1/[1 + (3/2)((n-1)/n)(S_t2/\Delta^2)]$.

Remark 2.5: If $S_t2 = 0$, then $E(\hat{\beta}_1 | Y_2 > c) = - \tau/\Delta$. That is, if all the subjects were exposed at the same time we would have an unbiased estimator of the decay rate. In this particular case there is no regression effect.

2.3 Some Applications

Example 1: As an application of Theorem 2.1, consider the simple unadjusted (for covariates) situation with two time points and suppose ΔT, the time from 1987 dioxin concentration until 1992 dioxin measurement, is the same constant for all subjects. Now, from the regression approach considered in Michalek et al. (1992), the uncorrected decay rate $\hat{\lambda}_u = D^-/\Delta T =$ the least square estimate of λ_u. Using currently available data, we have

$$\hat{\lambda}_u = 0.485929/5.113386 = 0.095031.$$

Then the uncorrected half-life $\hat{HL}$ is given by

$$\hat{HL}=(\ln(2)/0.095031=7.3 \text{ years}.$$

By selecting c_0 as the 0.40 quantile of the 1987 dioxin distribution, we get $k_0=0.64383$.

Using the formula given in Senn and Brown (1985), we have $\hat{\rho} = 0.892064$. The standard deviation of the logarithm of the 1987 concentration is $\hat{\sigma}=1.054023$.

Hence, the regression correction $= \{\hat{k}\ \hat{\sigma}\ (1-\hat{\rho})\}/\Delta T$.

$$= (0.64383)(1.054023)(1-0.892054)/5.113386$$
$$= 0.014389.$$

Therefore, the true decay rate is $\hat{\lambda}_t=\hat{\lambda}_u-\{\hat{k}\hat{\sigma}(1-\hat{\rho})\}/\Delta T$

$$= 0.095031-0.014389$$
$$= 0.080639767.$$

Thus, the corrected half-life is ($\ln(2)/0.080639767 = 8.6$ years.

Example 2: As an example of Theorem 2.2, we now find the corrected half-life using 1982 and 1987 dioxin concentrations. Using available data, the uncorrected decay rate is

$$\hat{\lambda}_u = 0.267815/5.189563 = 0.051485.$$

Now, using the formula given in Senn and Brown (1985), we have $\hat{\sigma}=1.054023$ and $\hat{\rho}=0.80787$.

Hence, the regression correction$=\{\hat{k}\ \hat{\sigma}(1-\hat{\rho})\}/\Delta T$.

$$= 0.025123821.$$

Therefore, the true decay rate is $\hat{\lambda}_t=\hat{\lambda}_u+\{\hat{k}\hat{\sigma}(1-\hat{\rho})\}/\Delta T$

$$= 0.051485 + 0.025123821$$
$$= 0.076608821.$$

Hence, the corrected half-life is ($\ln(2)/0.076608821$

$$=9.0 \text{ years}.$$

Example 3: As an example of regression effect on the slope, we now consider 240 Ranch Hand veterans with dioxin measurements taken in 1982, 1987 and 1992. In this case, $n=240$, $\Delta=5$ and $S_t2= 2.55$.

Therefore, the proportion of the regression effect on the slope is $1/[1+(3/2)((n-1)/n)(S_t2/\Delta^2)]=0.8695652$.

Taking 10 ppt as the cut point and using uncorrected decay rate

0.776, the corrected decay rate is found to be 0.08924. Therefore, the corrected half-life is (ln(2)/0.08924=7.77 years.

2.4 Discussion

Pirkle et al. (1989) estimated the half-life of dioxin concentration in Ranch Hand Veterans as 7.1 years using the standard half-life equation derived from first order kinetics. Michalek et al. (1992) used a repeated-measures analysis of a covariance model $y_{ij}=\beta_0+\tau_i+\beta_1 t_{ij}+\epsilon_{ij}$, and estimated the half-life using 1982 and 1987 dioxin concentration as 7.68 years. Recently, Wolfe et al. (1994) estimated the half-life as 11.6 years using a larger sample size (337). In our study we found an unbiased half-life estimate using the idea of regression towards the mean. First, we estimated the regression effect on the decay rate, and then we found the half-life using the corrected decay rate. Using 1982 and 1987 dioxin concentrations the corrected half-life was found to be 9.0 years, and using 1987 and 1992 concentration it was found to be 7.3 years. Using all three time points, that is, 1982, 1987 and 1992 dioxin concentrations, the half-life was estimated as 7.77 years. Here, the cut point is assumed to be 10 ppt. Moreover, we assumed that the data follows a multivariate normal distribution. Since the distribution of untransformed dioxin concentration is skewed the analysis could be carried out using a truncated bivariate exponential distribution. In our approach, we did not justify for selecting 10 ppt as the cut point. Perhaps, in future studies, we should be able to estimate the cut points using a truncated bivariate exponential distribution. Furthermore, we haven't looked at the robustness of the estimators. This would be an interesting study in the future.

Acknowledgement

References:

Anscombe, F. J. (1953). Sequential estimation. J. Roy. Statist. Soc. Ser. B, 151-29.

Bar-Lev, S. K. and Boukai, B. (1985). Minimum variance unbiased estimation for the families of distributions involving two truncations parameters. J. Stat. Plan. and Inf. 12, 379-384.

Csorgo, M. (1968). On the strong law of large numbers and the central limit theorem for martingales. Trans. Ann. Math. Soc. 131, 259-275.

Davis, R. C. (1951). On minimum variance in nonregular estimation, Ann. Math. Statist. 22, 43-57.

Ferentinos, K. K. (1987). Shortest confidence intervals and UMVU estimators for families of distributions involving truncation parameters, Metrika 34, 341-359.

Guenther, W. C. (1978). Some easily found minimum variance unbiased estimators, Amer. Statist. 33, 29-34.

James, K. E. (1973). Regression toward the mean in uncontrolled clinical studies, Biometrics 29, 21-13.

Krishnamoorthy, K., Rohatgi, V. K. and Blass, Josef (1989). Unbiased estimation of type II censored samples from a one-truncation parameter density, Commun. Statist.-Theory and Methods 18(3), 1023-1030.

Selvavel, K. (1989). Unbiased estimation in sampling from one-truncation parameter families when both samples are type II censored. Commun. Statist.-Theory and Methods 18(9), 3519-3531.

Senn, S. J. and Brown, R. A. (1985). Estimating treatment effects in clinical trials subject to regression to the mean, Biometrics 441, 555-560.

Wald, A. (1947). Sequential analysis, Wiley, New York.

Woodroofe, M. (1982). Nonlinear renewal theory in sequential analysis. Society for Industrial and Applied Mathematics. Philadelphia.

MULTISITE OPTICAL RECORDING OF EVOKED ACTIVITY
IN MOUSE SUPRACHIASMATIC NUCLEI

David M. Senseman

Associate Professor

Division of Life Sciences

The University of Texas at San Antonio

6900 N. Loop 1604 W.

San Antonio, TX 78249

MULTISITE OPTICAL RECORDING OF EVOKED ACTIVITY IN MOUSE SUPRACHIASMATIC NUCLEI

David M. Senseman

Associate Professor

Division of Life Sciences

The University of Texas at San Antonio

Abstract

Fast multisite optical recording techniques were used to monitor activity in mouse suprachiasmatic nuclei (SCN) evoked by direct electrical stimulation of the SCN tissue with a focal electrode. Experiments were performed on coronal brain slices (400-500 μm thick) obtained from an inbred mouse strain (C57BL/6NHsd). Under normal conditions, the amplitude of optically recorded signals evoked by focal electrical stimulation was found to be quite small. Signal size was not significantly improved by $GABA_A$ receptor blockade with picrotoxin. A large increase in signal size was observed, however, following bath application of the potassium channel blocker, 4-aminopyridine (4-AP). In slices treated with 4-AP, focal stimulation generated electrical activity that propagated medially into the contralateral SCN and dorsally into an area ventral to the paraventricular nucleus of the hypothalamus (PVN) known as the subperiventricular region. Propagated activity could be abolished by the voltage-sensitive sodium channel blocker, tetrodotoxin, indicating that is was being conducted by nerve action potentials in axonal fibers. Although it seems likely that most, if not all, of these axonal fibers arose from SCN efferents, activity in retinohypothalamic fibers could not be ruled out in the current series of experiments. The significance of these finding for future investigations of the SCN efferent system is discussed.

MULTISITE OPTICAL RECORDING OF EVOKED ACTIVITY IN MOUSE SUPRACHIASMATIC NUCLEI

David M. Senseman

INTRODUCTION

Before the role of the suprachiasmatic nucleus (SCN) in circadian biology can be determined, three fundamental questions must be answered (Watts, 1991). First, how does the SCN generate a circadian signal? Second, how is this signal entrained by light? Third, how is this circadian signal transferred into the systems of the brain to control rhythmical processes? Compared to the first and second questions, relatively little research effort has been directed to resolving the third question, the nature of the SCN's efferent output (cf. Watts, 1991; Joosse et al., 1992).

Most of what is currently known about the efferent output of the SCN is anatomical in nature derived from histological studies of the rat (van den Pol, 1980; 1991; van den Pol and Tsujimoto, 1985; Berk and Finkelstein, 1981; Stephan et al., 1981; Swanson and Cowan, 1975; Watts and Swanson, 1987; Watts et al., 1987) and, to a lesser extent, the golden hamster (Kalsbeek et al., 1993). The major projection of the SCN leaves the nucleus and runs dorsocaudally toward the ventral border of the paraventricular nucleus of the hypothalamus (PVH) (Watts, 1991). Most of the fibers terminate in area ventral to the PVH known as the subparaventricular zone.

Recently, Senseman and Rea (1994) showed that voltage-sensitive dyes could be used to monitor electrical activity in the hamster SCN evoked by electrical stimulation of the retinohypothalamic tract. Using a silicon photodiode array as their imaging device, these authors were able to monitor activity, simultaneously, in up to 464 contiguous anatomical regions. The overall objective of the current series of experiments was to see if a similar experimental approach could be used to monitor the efferent output of the SCN evoked by focal electrical stimulation.

METHODOLGY

Experiments were performed on an inbred mouse strain (C57BL/6NHsd) obtained from Harlan Spague-Dawley Inc. Animals were housed in an IACUC approved central animal facility on a strictly enforced 14:10 LD cycle (lights on at 2300 h) for at least 2 weeks prior to use.

Hypothalamic brain slices were prepared following the procedure of Hatton et al. [11]. Prior to decapitation, a deep level of anesthesia was induced with Metofane. The brain was then rapidly removed from the cranium and placed in chilled (4° C) mammalian Ringer's solution (NaCl 125 mM, KCl 5 mM, $MgSO_4$ 2 mM, $CaCl_2$ 3 mM, NaH_2PO_4 25.5 mM, D-glucose 10 mM) for 30 s to reduce metabolic activity. A block brain tissue containing the SCN's, optic nerves and chiasm was prepared and mounted on the stage of an oscillating tissue slicer (Fredrick Haer model OTS-3000) using cyanoacrylate adhesive. A series of 400-500 μm thick slices were cut in the coronal plane and placed in an moist incubation chamber supplied with continuously flowing 95% O_2- 5% CO_2. Following a 30 min recovery period, the slice was stained for 60 min with the voltage-sensitive dye RH155 (Nippon Kankoh-Shikiso Kenykyusho Co., Okayama, Japan). After removing excess dye with fresh mammalian Ringer's, slices were allow to recover an additional 30 min in the incubation chamber before the first experimental trial.

For optical recording, a stained slice was removed from the incubation chamber and secured to the Sylgard-coated glass bottom of a recording chamber by means of miniature tungsten staples. In some experiments, a suction electrode was attached to the optic chiasm. The chamber was mounted on the stage of a large binocular microscope (Zeiss UEM) and the preparation perfused with 10 ml/min oxygenated mammalian Ringer's. Perfusion was maintained throughout the duration of the experiment except for brief periods (20-30 sec) during the actual acquisition of the optical signals. This interruption of flow was necessary to minimize contamination of the optical signals by vibrational noise.

Fig. 1 shows a semi-diagrammatic representation of the experimental setup. The preparation was trans-illuminated with light from a 100-watt tungsten-halogen lamp driven by a stable DC power supply. The collimated light was passed through a narrow bandpass interference filter (7300 +400 nm) before being focused on the preparation by means of the microscope's substage condenser. A 16X (0.5 NA) water immersion objective (Zeiss) collected the transmitted light and formed a real image of the preparation on the surface of a 464-element silicon photodiode array (Centronic Inc., model MD-464).

Optical Arrangement
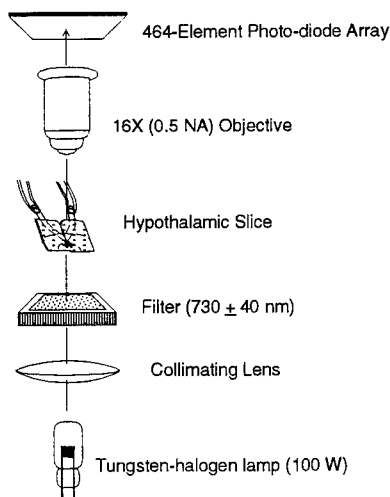
Fig. 1. Schematic drawing of the experimental apparatus. A 24 x 24 silicon photodiode array was used for high-speed optical recording (1ms/frame) of voltage-sensitive signals from the experimental preparation (brain). Photo-currents from the 464 central array elements were passed in parallel to an amplifier bank (not shown) and then to a 512 channel digital-to-analog converter system (not shown) before processing on a 32-bit Motorola MC68030-based microcomputer system (data acquisition system). A Hammamatsu C2400 video camera system was used at the end of an experiment to record a high-resolution video image of the experimental preparation. The video image was stored on a Panasonic video disc recorder prior to digitization and processing on a Imaging Technology Series 151 image processor system. A Silicon Graphics Inc. 4D/RPC color graphics workstation produced the animated PAM displays from the combined optical recording data and the enhanced image data.

The photocurrent output from each diode element was passed in parallel to a 2-stage amplifier system that amplified and low- and high-pass filtered the signals before being digitized (16-bit resolution) and stored on a MC68030-based computer system (Motorola MVME-147). Details about the amplification and digitization subsystems have been published elsewhere [12]. At the end of the experiment, the photodiode array was replaced with a conventional video camera (Hammamatsu model C2400) and a high-resolution (512 x 512 pixels) grey-scale (8-bit) image of the preparation acquired using a low power (4X) microscope objective. The image was stored on a optical disc memory recorder (Panasonic model TQ-2025F) and later digitized with a image processor system (Imaging Technologies Inc. series 151). Digitized video and photodiode array information was transferred to a 64-bit MIPS R4000-based UNIX workstation (Silicon Graphics Inc. model 4D/RPC "Indigo") for data display and analysis as previously described [13].

RESULTS

Figure 2 shows an example of a mouse coronal brain slice containing the optic chiasm, SCN and surrounding hypothalamus. The approximate boarders of the SCN are indicated as well as the location of the focal stimulating electrode. The optic chiasm is visible at the bottom of the figure. Grid lines superimposed on the slice show the areas monitored by each element in the 464-element silicon photodiode array detector.
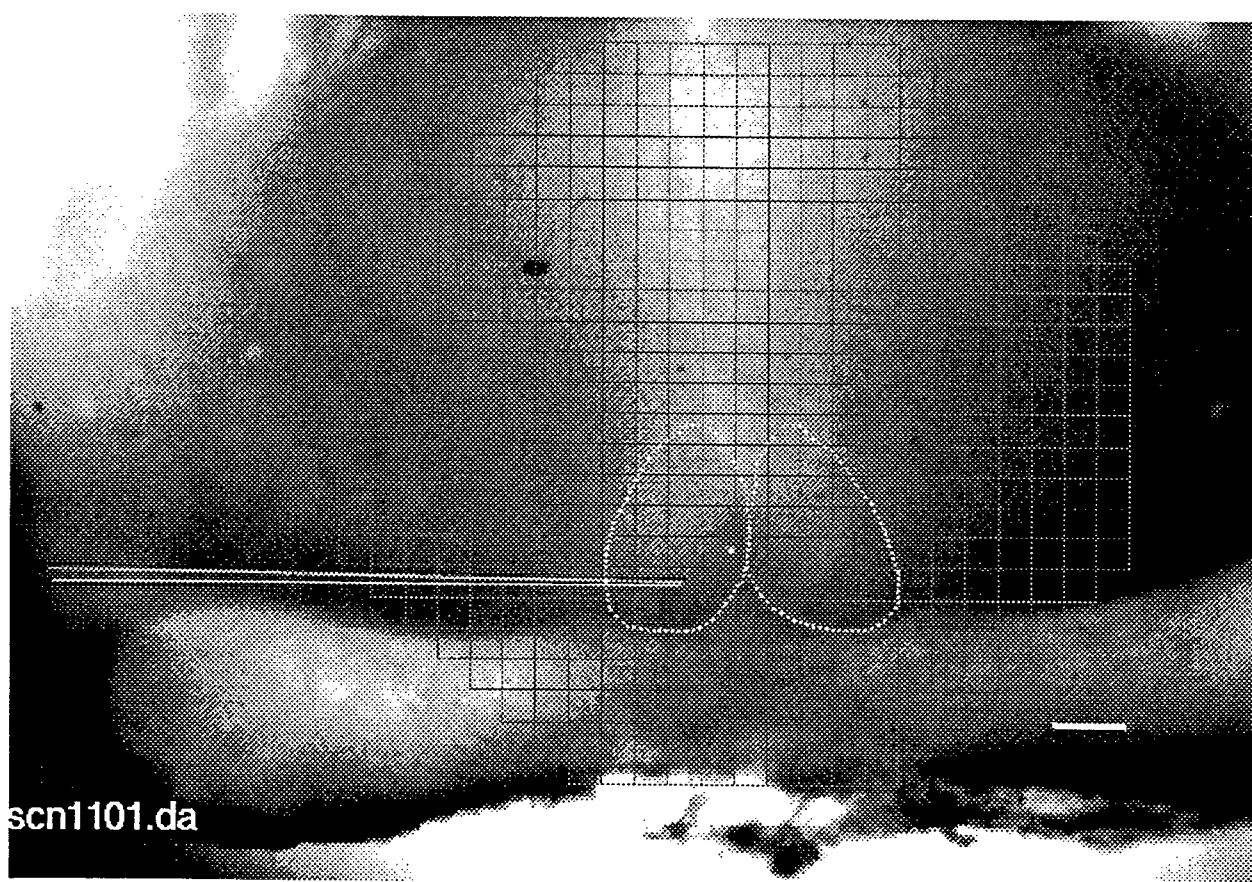
Fig. 2. Coronal slice preparation of the mouse brain used for optical recording. The optic chiasm is visible at the bottom of the slice. The approximate boarders of two SCN nuclei are indicated. See text for further details. Calibration bar is 100 μm.

Under normal conditions, focal electrical stimulation of the SCN produced relatively small optical signals that attenuated rapidly with distance from the site of stimulation (Fig. 3). We considered the possibility that such small signals might be due to tonic GABAergic inhibition of the intrinsic SCN neural network. Bath application of the $GABA_A$ blocker, picrotoxin (100 μM),

produced a small, but measurable increase in the amplitude of evoked signals recorded 200-300 μm distal to the site of stimulation as well as near its origin (Fig. 3). A small increase in signal size was also observed after blockade of GABA$_B$ receptors with CGP 55845A (50 μM) (data not shown). In contrast, very large optical signals were recorded near the origin of stimulation and in more distal locations following bath application of the potassium channel blocker, 4-aminopyridine (4-AP).
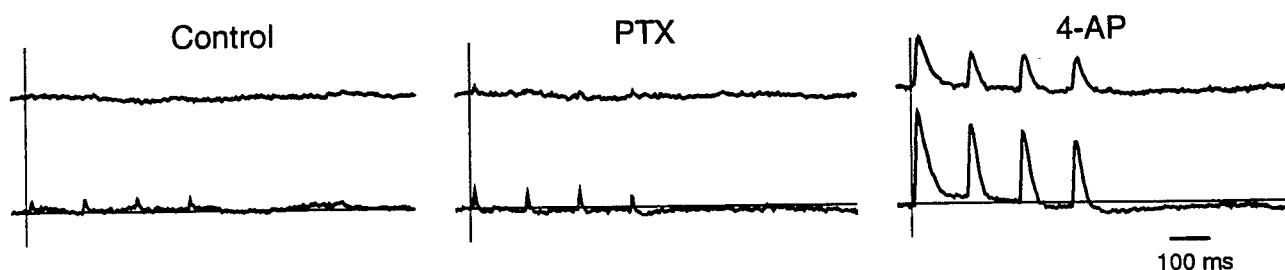
Fig. 3. The effects of picrotoxin (PTX) and 4-aminopyridine (4-AP) on the focally evoked activity. Lower trace in each pair was recorded near the site of focal stimulation, the upper trace was recorded 200 μm distal to the origin of stimulation. Focal stimulation was a train of four shocks (5.0 mA, 1 ms) delivered at 4 Hz. Vertical line indicates stimulus onset for the initial shock. Concentration for both PTX and 4-AP was 100 μM.

We considered the possibility that the optical signals we recorded following 4-AP treatment might reflect simply an enhancement of passive electrotonus. Current injected into the SCN could depolarize directly the membranes of neurons and/or glial elements which would spread passively from the site of stimulation. This possibility seemed unlikely since reversing the polarity of the stimulating current decreased did not invert the signal as would be expected for electrotonus. Nevertheless, we wanted to establish clearly the active, regenerative nature of the evoked responses, particularly those recorded from more distal locations. We used voltage-dependent sodium channels blocker, tetrodotoxin (TTX), to eliminate regenerative neural activity. TTX-poisoning would be expected to suppress the propagation of sodium-based action potentials in axonal fibers but have little effect, if any, on the passive spread of electrotonic current. We found that TTX (1 μM) completely eliminated all focally-evoked activity in regions > 100 μm from the site of stimulation (Fig. 4). At, and immediately surrounding the stimulation site, TTX produced a large reduction in signal amplitude leaving a small, residual response. We believe this residual

response reflects the direct electronic depolarization of neurons and/or glial surrounding the electrode tip. Alternatively, the residual depolarization might reflect an inward calcium current through voltage-activated calcium channels which are not blocked by TTX. Further experiments are required to settle this point.

Control   **TTX**

100 ms
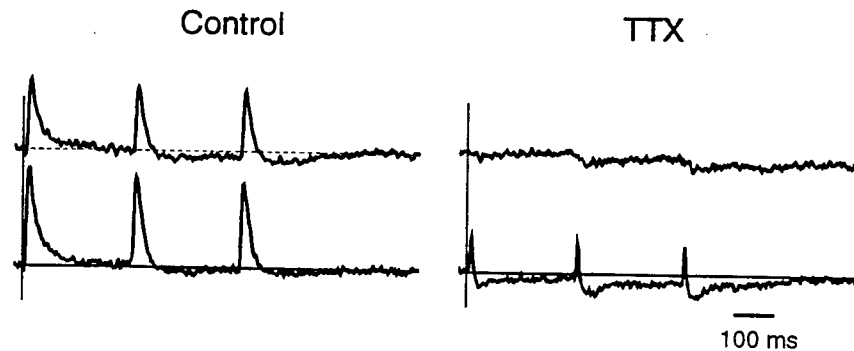
Fig. 4. The effects of tetrodotoxin (TTX) on the focally evoked activity. Lower trace in each pair was recorded near the site of focal stimulation, the upper trace was recorded 250 μm distal to the origin of stimulation. Focal stimulation was a train of four shocks (5.0 mA, 1 ms) delivered at 3 Hz. Vertical line indicates stimulus onset for the initial shock. Concentration of TTX was 1 μM.

A major objective of these experiments was to use the multisite recording capability of our data acquisition system to follow the propagation of evoked activity from its site of origin into both ipsilateral and contralateral hypothalamic regions. We found that focally stimulation of the SCN in slices treated with 4-AP, produced a very large depolarization within the stimulated nucleus that could be observe to propagate dorsally into the subparaventricular zone and medially into the contralateral SCN (Fig. 5).

Three-dimensional surface plots of the evoked responses were prepared in order to visualize more clearly that relative amplitude of the responses recorded within the SCN with the amplitude of the responses recorded in surrounding hypothalamic areas.Figure 6 shows a three-dimensional contour plot of the data presented in Fig. 5. It can be seen that signal amplitude attenuates rapidly outside the ipsilateral SCN.
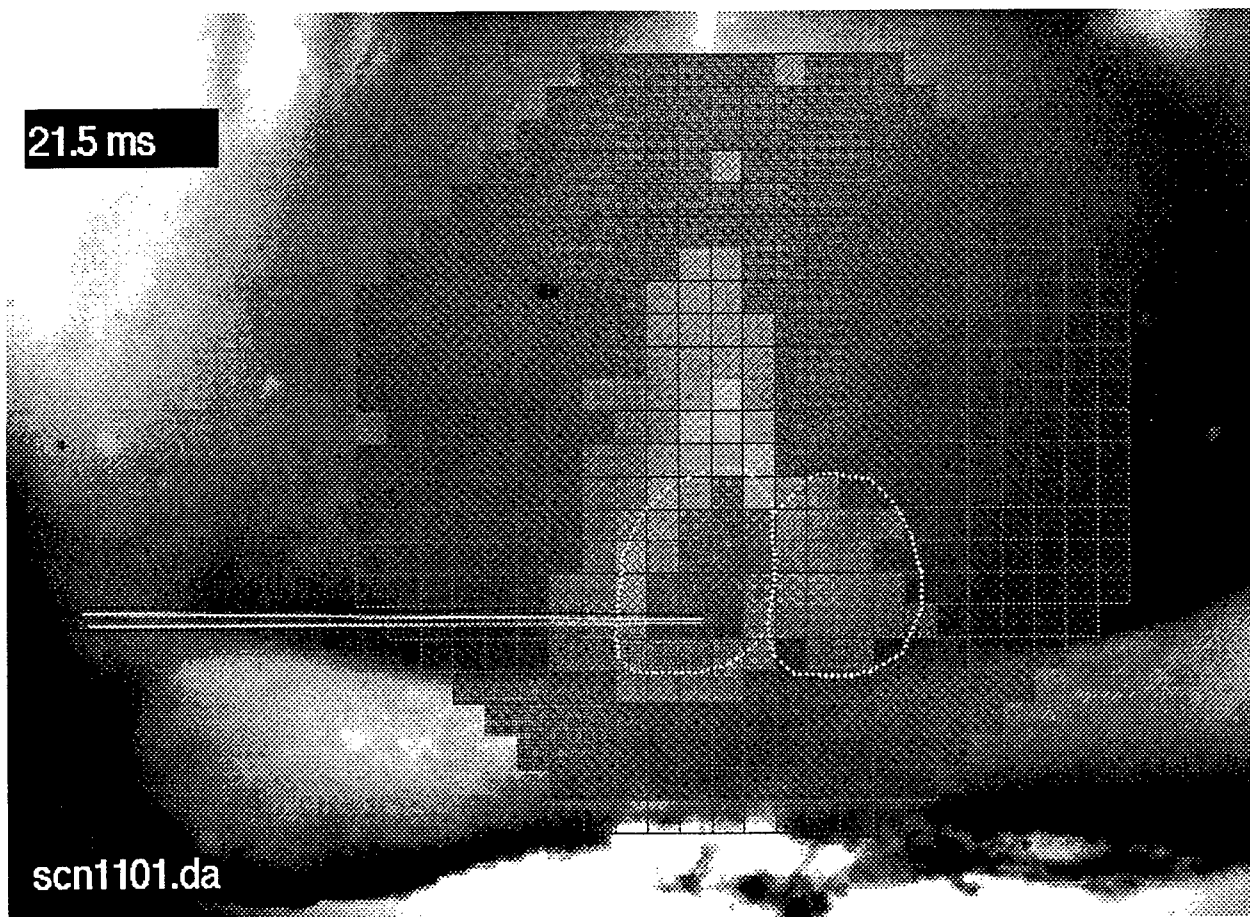
Fig. 5. Pseudocolor activity map showing propagation of neural activity from the SCN into the ipsilateral subperiventicular zone and into the contralateral SCN. The approximate boarders of two SCN nuclei are indicated. See text for further details. Calibration bar is 100 μm.
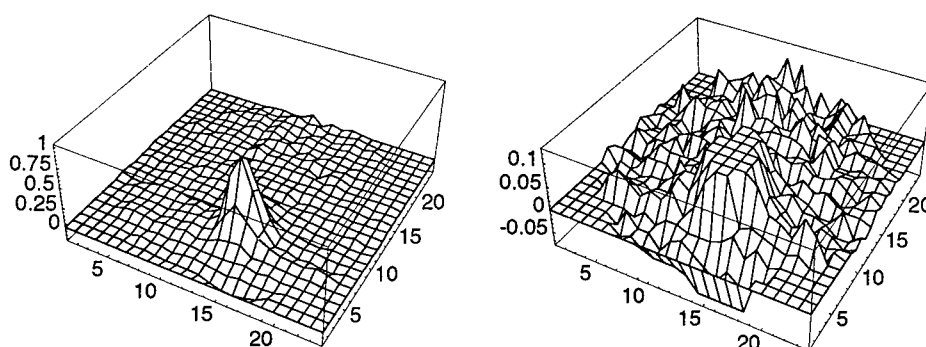
Fig. 6. Three-dimensional contour plots of the data presented in Fig. 5. The same data is presented in both plots; the vertical scale on the right hand plot has been increased to show the pattern of activity in the areas surrounding the SCN. Note how rapidly signal amplitude falls off beyond the boarders of the SCN.

DISCUSSION

A major objective of the current series of experiments was to determine the feasibility of using multisite optical recording techniques to following SCN efferent activity to its neuronal targets. While a number of technical problems still need to be resolved, the preliminary results obtained during the 1994 Summer Faculty program at Brooks AFB, are encouraging. We have established that relatively large optical signals can be recorded from coronal brain slices obtained from a genetically inbred mouse strain following treatment with 4-AP. Since these signals are abolished by TTX, we believe they represent propagated neural activity and not direct electrotonic depolarization of neuronal and/or glial membranes. Neural activity evoked in one SCN can be seen to propagate dorsally, into a region ventral to the PVH, known as the subperiventricular zone and medially into the contralateral SCN. This result directly confirms previous histological and immunohistochemical studies (c.f. Watts, 1991, van den Pol, 1991).

One unresolved issue is the extent to which propagated activity produced by focal stimulation is carried by axonal fibers of SCN neurons or by retinohypothalamic fibers of passage. This issue could be resolved by repeated the above experiments in animals in which the optic fibers have degenerated in response to enucleation.

LITERATURE CITED

Berk, M.L. and Finkelstein, J.A. 1981. An autoradiographic determination of the efferent projections of the suprachiasmatic nucleus of the hypothalamus. Brain Res. **226**: 1-13.

Hatton, G.I., Doran, A.D., Salm, A.K. and Tweedle, C.D. 1980. Brain slice preparations: hypothalamus. Brain Res. Bull. **5**: 405-414.

Kalsbeek, A., Teclemariam-Mesbah, R., and Pevet, P. 1993. Efferent projections of the suprachiasmatic nucleus in the golden hamster *(Mesocricetus auratus)*. J. Comp. Neurol. **332**: 293-314.

Kalsbeek, A. and Buijs, R.M. 1992. Peptidergic transmitters of the suprachiasmatic nuclei and the control of circadian rhythmicity. In: *Progress in Brain Research Vol. 92*: ed. J. Joosse, R.M. Buijs and F.J.H. Tiders. Elsevier, pp. 321-333.

Senseman, D.M. and Rea, M.A. 1994. Fast multisite optical recording of mono- and poly-synaptic activity in the hamster suprachiasmatic nucleus evoked by retinohypothalamic tract stimulation. NeuroImage (in press).

Senseman, D., Vasquez, S. and Nash, P.L. 1991. Animated pseudocolor activity maps (PAM's): Scientific visualization of brain electrical activity. In: *Chemosensory Information Processing.* ed D. Schilds, NATO ASI Series Vol. H 39, Springer-Verlag, Berlin, 1990, pp. 329-347.

Swanson, L.W. and Cowan, W.M. 1975. The efferent projections of the suprachiasmatic nucleus of the hypothalamus. J. Comp. Neurol. **160**: 621-656.

van den Pol, A.N. 1980. The hypothalamic suprachiasmatic nucleus of the rat intrinsic anatomy. J. Comp. Neurol. **191**: 661-702.

van den Pol, A.N. and Tsujimoto, K.L., 1985. Neurotransmitters of the hypothalamic suprachiasmatic nucleus: immunocytochemical analysis of 25 neuronal antigens. Neuroscience **15**: 1049-1086.

van den Pol, A.N. 1991. The suprachiasmatic nucleus: Morphological and cytochemical substrates for cellular interaction. In: *The Mind's Clock.* ed. D.C. Klein, R.Y.Moore and S.M. Reppert. Oxford Univ. Press, Oxford. pp 17-50.

Watts, A.G. 1991. The efferent projections of the suprachiasmatic nucleus: anatomical insights into the control of circadian rhythms. In: *The Mind's Clock.* ed. D.C. Klein, R.Y. Moore and S.M. Reppert. Oxford Univ. Press, Oxford. pp 77-106.

Watts, A.G. and Swanson, L.W. and Sanchez-Watts, G. 1987. Efferent projections of the suprachiasmatic nucleus: I. Studies using anterograde tansport of *Phaseolus vulgaris* leucoagglutinin in the rat. J. Comp. Neurol. **258**: 204--229.

Watts, A.G. and Swanson, L.W., 1987. Efferent projections of the suprachiasmatic nucleus: II. Studies using retrograde transport of fluorescent dyes and simultaneous peptide immunohistochemistry in the rat. J. Comp. Neurol. **258**: 230-252.

LINKING LABORATORY RESEARCH AND FIELD APPLICATIONS
OF COMPLEX SKILL ACQUISITION

Wayne Shebilske
Professor
Department of Psychology

Texas A&M University
College Station, TX 77843-4235

Final Report for:
Summer Research Program
Armstrong Laboratory

LINKING LABORATORY RESEARCH AND FIELD APPLICATIONS OF COMPLEX SKILL ACQUISITION

Wayne Shebilske
Professor
Department of Psychology
Texas A&M University

Abstract

Data were analyzed and articles were prepared for three laboratory studies that were designed to represent complex skills learned in military and industrial settings. Effects of competition, spacing of practice, and gender, were investigated. In addition, a theoretical article was prepared to discuss the implication of this and other recent research for expanding the role of skill acquisition specialists in the evolution of effective human-interfaced systems.

LINKING LABORATORY RESEARCH AND FIELD APPLICATIONS

OF COMPLEX SKILL ACQUISITION

Wayne Shebilske

Introduction

The gap between laboratory research on skill acquisition and applications in military and industrial settings is often wide because many laboratory paradigms are not representative of training in applied settings.   Field research is more representative, but it often lacks the control necessary for precise analysis.   Research that bridges this gap must faithfully represent the cognitive processes that underlie training in the field, and it must enable rigorous analysis of those processes. Research at the TRAIN (Training Research for Automated INstruction) Laboratory is doing both.

The present research employed Space Fortress, a video game-like research tool. It's representativeness of training in applied settings is low with respect to physical conditions but high with respect to cognitive processes (Donchin, 1989; Mane & Donchin, 1989).   Past research employing Space Fortress at the TRAIN laboratory established a protocol for teaching complex skills in small groups (Shebilske, Regian, Arthur & Jordan, 1992). Research on the protocol isolated and analyzed specific factors , such as observational learning, that makes group protocols very efficient (Shebilske & Regian, 1992).   The research also inspired protocol designs at Aer Lingus pilot training center in Dublin (Johnston,

Regian, & Shebilske, 1994). Thus, a bridge was anchored on both sides of the gap between research and applications. That bridge is now being used to convey ideas back and forth between laboratory research and field applications. The present research expands the bridge into the areas of competition, spacing of practice, and gender. Each of these areas has the potential for important applications based on rigorous empirical and theoretical foundations.

General Methodology

Subjects

Subjects were recruited by several temporary agencies and paid about $5 for each hour of participation in the study. Upon recruitment, subjects were informed that they would be performing several video-game like computer tasks and completing measures of individual differences and characteristics. All subjects were between 18 and 27 years of age. Subjects were screened and not allowed to participate in the study if they reported currently playing video games more than 20 hrs per week or if they failed to achieve a minimum score of 780 on at least one of three, 2-min games of an Aiming screening task (Shebilske, Regian, Arthur, & Jordan, 1992; Mane & Donchin, 1989).

Procedures

Trainees in the present studies practiced Space Fortress for 10 sessions. Every session had 8 practice games of 3 minutes each and 2 test games of 3 minutes each. Independent variables were manipulated by randomly assigning subjects to different practice

conditions; and dependent variables were measured during tests, which were identical for all trainees.

The major game components were a computer screen, a left-handed response buttons, and a right-handed joystick. A fortress occupied the screen's center, and two concentric hexagons surrounded it. An information panel ran across the screen's bottom. The subject controlled a space ship's flight path with the joystick and shot missiles with a trigger on the joystick. Fortress vulnerability was given in the information panel and was changed with each missile hit. Mines also flew in the space surrounding the fortress. They were either friends or foes, and they were identified by the mine indicator in the information panel. Foe mines require subjects to push the Identify Friend or Foe (IFF) button at the appropriate time. Symbols appeared below the fortress to indicate opportunities for bonuses, which were collected by pushing either the points (P) or missile (M) button at the appropriate time. The information panel also displayed the number of missiles available to the ship, a score determined by destroying and not being destroyed, a score based on ship velocity, a score based on ship control, and a score based on the speed of handling mines. After each game, the screen displayed a total score, which was a composite of the others. The total score was the main dependent variable in three separate studies on the effects of competition, spacing of practice, and gender.

Study 1 : Competition

Competition has been used to stimulate computer-based

Instruction (CBI) in schools, business, and athletics. However, despite the ubiquitousness of competition effects and theoretical accounts of them, much is unknown about the effect and how to apply it. Understanding that might guide applications in CBI is obscured in part by a gap between theory and practice. The empirical and theoretical foundation is based mostly on tasks that are unrepresentative of complex skill acquisition in CBI. Study 1 attempts to bridge the gap by analyzing the effect of competition on the learning of Space Fortress.

<center>Method</center>

Competition Manipulation

Competition manipulations were administered 2 times during the experiment, once at the beginning of Day 1 and once at the beginning of Day 2. Through these manipulations, subjects were placed in either a competitive (Competition) or non-competitive (No-Competition) performance environment. In the competitive environment, subjects read instructions informing them that they would be competing individually against all other subjects in the study and that the subjects with the first, second, and third highest scores on the SF task (based on the average of the 2 test games at the end of each acquisition session) would win $100, $60, and $40, respectively, in addition to their normal pay. In the non-competitive environment, subjects read instructions informing them that they would be entered into a $100 lottery if they completed the study.

Early Competition. Subjects in this condition were given

the Competition manipulation on Day 1 and the No-Competition manipulation on Day 2.

No Early Competition. Subjects in this condition served as controls for subjects in the Early Competition condition and were run at the same time. These subjects received the No-Competition manipulation on Day 1 and Day 2.

Late Competition. Subjects assigned to this condition were given the No-Competition manipulation on Day 1 and the Competition manipulation on Day 2.

No Late Competition. The subjects in this condition served as controls for subjects in the Late Competition condition and received the No-Competition manipulation on Day 1 and Day 2.

Results

The baseline score correlated with performance on Session 10, the final acquisition session, (r=.45, p<.001), and on Session 11, the retention test (r=.45, p<.001). Therefore, the baseline was used as a covariate in all analyses of covariance (ANCOVA). The baseline scores for the four competition conditions did not differ significantly from one another ($F(3,76)$<1). Furthermore, although the No Early Competition condition performed substantially better than the No Late Competition condition, this difference did not reach statistical significance ($F(1,37)$=2.59, p>.05). Nevertheless, we followed our a priori plan of comparing the Early Competition condition with the No Early Competition condition and comparing the Late Competition condition with the No Late Competition condition.

The Early Competition condition did not differ from the No Early Competition condition. A repeated measures ANCOVA indicated that improvement across sessions was significant for both conditions ($F(10,370)=3.89$, $p<.001$), but that the two conditions did not differ overall ($F(1,37)<1$), on Session 10 ($F(1,37)<1$), or on Session 11 ($F(1,37)<1$). In contrast, the Late Competition condition performed better than the No Late Competition condition. Improvement across sessions was significant ($F(10,370)=9.48$, $p<.001$), and the Late Competition condition performed better than the No Late Competition condition overall ($F(1,37)=5.06$, $p<.05$). In addition, the Late Competition condition performed better than the No Late Competition condition on Session 10 ($F(1,37)=4.89$, $p<.05$) and on Session 11 ($F(1,37)=5.13$, $p<.05$).

Discussion

The results are consistent with social learning theory according to which the arousal of competition facilitates dominant responses. When competition is introduced early in training, the dominant responses tend to be wrong; when competition is introduced later in training, the dominant responses tend to be correct. For example, beginners tend to grip the joystick firmly and move it from one extreme position to another. More experienced trainees tend to grip the joystick lightly and tap it gently from one position to another. The arousal of competition that is introduced early in training facilitates the dominant, inappropriate heavy-handed responses. The arousal of competition

that is introduced later in training facilitates the dominant, appropriate light-handed responses.

Study 2: Spacing of Practice

Most previous research and theories about massed practice concern massing over a relatively short time period (e.g., Glenberg, 1979). Spacing within simple tasks was manipulated on a micro scale, in terms of seconds and minutes. The question arises whether these theories apply to massing of complex tasks on a macro scale, defined in terms of hours and days. It is too great an extrapolation to assume generalization from seconds and minutes to hours and days. Variables exist that apply to distribution of practice over days but not to distributing within a one-hour session. For example, fatigue is a factor that may apply to practice over hours and days, but it seems unlikely to effect practice over seconds and minutes.

Even variables that logically apply to both paradigms may display differential interactions with other variables. One such variable is confidence. A study by Green (1991) provides an example. In his study, massed subjects employed less overt rehearsal than distributed subjects. Green attributed this to a metacognitive error. The massed subjects reported more confidence in their performance than the distributed subjects, despite the fact that they were actually performing worse. A second factor that might operate differently in micro and macro situations is retrieval from long-term memory. Spacing of practice forces a person to reconstruct a representation from

long-term memory (Schmidt & Bjork, 1992). According to this formulation, massed practice is not as effective because it only requires the use of short-term memory during practice. While confidence and long-term retrieval could potentially mediate the effect of practice distribution at both the micro and the macro level, they probably interact differently depending on the time frame.

Another difference between micro and macro studies relates to the definition of spacing. In micro studies, distribution of practice is defined in terms of the intertrial interval, while macro studies typically examine the amount practiced in a day. Thus, even if the distributed practice effect is replicated at the macro level, the theories to explain it may be modifications or entirely different from those used in conventional paradigms.

Method

Subjects played 4 baseline games, followed by 10 sessions of 8 practice games and 2 test games. After Session 1, two consecutive reaction time tests (USRT) were administered. These served to familiarize subjects with the task. This measure was taken 5 more times, after Space Fortress Sessions 2, 4, 6, 8, and 10. Prior to Space Fortress Session 2, a confidence questionnaire was administered. This measure was taken again before Sessions 4, 6, 8, and 10.

On the final day of practice, after Session 10 and the USRT, subjects performed a battery of retention, secondary, and transfer tests. Session 11 consisted of two, 3-minute test games

on a secondary tapping task. Session 12 was two, 3-minute retention games of Space Fortress. Subjects performed the two, 3-minute test games on a keyboard transfer task. And Session 14 was a repeat of the 2 retention games.

One week after the final day of practice, subjects returned to complete an exit survey and be paid. They were unaware that they would perform additional retention, secondary, and transfer tests. The tests during this "Pay" Session were identical to Sessions 11-14, with the exception that two additional retention games preceded the secondary task. After performing the tasks, subjects were paid and debriefed. After an analysis of the test games, bonuses were paid to subjects that had the highest 3 scores.

Results

To determine if the baseline games were an appropriate covariate, the correlation between these games and Session 10 test games was examined. This analysis collapsed across all conditions. Results indicated a moderate correlation ($r = .60$, $p < .05$) between baseline and Session 10 test games, suggesting that the baseline games were an appropriate covariate for utilizing repeated measures analyses of covariance.

An analysis of variance (ANOVA) was conducted on the average score of the four baseline games across all conditions. There was no main effect for condition ($p > .05$). This indicates that the random assignment of subjects to conditions was successful.

Manipulation checks across all conditions and sessions were

performed on the extraneous variables. Analyses of variance showed no significant effect of practice station or practice time on condition ($p > .05$). Therefore the remaining analyses collapsed over these two extraneous variables.

Directional repeated measures analyses of covariance (ANCOVA) were conducted with the baseline games serving as the covariate. Overall analyses revealed main effects for session ($F (9, 378) = 16.61, p < .001$) and spacing ($F (1, 42) = 4.23, p < .05$) as shown in Figure 1. Distributed subjects had a higher average score than the massed subjects. The analysis also showed an interaction between session and spacing ($F (9, 378) = 3.71, p < .01$), suggesting that the distributed group learned at a faster rate than the massed group.

Analyses of variance (ANOVA) were conducted on all retention, secondary, and transfer sessions. Both immediate retention tests proved significantly different, ($F (1, 42) = 8.03, p < .01$ and $F (1, 42) = 7.01, p < .05$, respectively). Distributed scores were higher than massed scores. The one-week secondary tests ($F (1, 42) = 10.39, p < .01$) indicated significant performance differences between groups. Distributed subjects performed better than massed subjects. No significant differences were obtained for either the immediate or the one-week transfer tests.

USRT results indicated no significant differences between groups. Reactions times were comparable for the massed and the distributed conditions.

No significant results were found for the majority of the items on the confidence questionnaire. However, distributed group confidence ratings were significantly higher than massed group ratings when asked if they thought their final test game scores would be high relative to other participants in the study, ($F (1, 38) = 5.93$, $p < .05$). Similarly, distributed subjects had higher rating scores than massed subjects when asked if they thought they would improve on the upcoming session, ($F (1,38) = 6.78$, $p < .05$). In contrast, massed group ratings were higher than distributed group ratings when asked if they were confident that they understood the optimal rules and strategies of the game, ($F (1, 38) = 7.75$, $p < .01$).

In the questionnaire, subjects were also asked if they were sleepy, if they were motivated to play the game, and if they were alert. Both groups indicated low alertness and motivation with no significant differences between groups.

Discussion

The short-term goal of evaluating the effects of intersession practice distribution on acquisition and retention of a complex task was met. The results contrast with previous studies that found only retention differences (Smith & Rothkoph, 1984) or only acquisition differences (Adams, 1987) between distributed and massed trainees. In the present study, distributed practice increased performance during acquisition, and the performance difference was maintained on a 1-week retention test. These results suggest that intersession

distribution of practice yields a true learning advantage for the distributed group as opposed to a temporary performance decrement for the massed group.

Laboratory conditions enabled detailed analyses of many independent and dependent variables, which shed light on potential causes of the distributed effect in complex skill acquisition. Fatigue, confidence, and motivation were ruled out as possible causes, which is an important step in the analytic process of elimination in understanding and remediating suppression caused by massing complex tasks within days.

Study 3: Gender

The largest difference that we have obtained so far in Space Fortress performance is between males and females. A step toward understanding these differences, the present study will provide a preliminary description of differences between 47 females and 104 males.

Method

The methods were identical to the control conditions in the previous studies except that in addition to 104 males, the sample included 47 females.

Results

Total scores on baseline tests were -1342 for males and -2236 for females ($\underline{F}(1,149)= 38.30$, $\underline{p}< .001$). Total scores on session 10 were 2200 for males and 517 for females ($\underline{F}(1,148)=7.19$), $\underline{p},.01$). Total scores on session 11, a retention test, were 2211 for males and 417 for females ($\underline{F}(1,148)=10.14$,

$p<.005$). Cohen's d for these gender differences indicated that males performed 0.973 standard deviations better than females on the baseline, 1.5 standard deviations better on Session 10 and 1.007 standard deviations better on the retention test.

Discussion

Understanding these gender differences has important theoretical implications. Shute (1983) found that performance on spatial skills is predicted by levels of male sex hormones, androgens. Men who were on the low end of the normal male range for androgens were better than males at the high end. In contrast, women who were at the high end of the normal female range of androgens were better than women at the low end. This pattern suggest that optimal androgen levels for spatial skills might be in a moderate range that overlaps the high end of the female range and the low end of the male range. This conclusion has been supported recently (Gouchie and Kimura, 1991; Kimura, 1992). The TRAIN laboratory is preparing to follow up this exciting area of research by replicating the present study along with measurements of hormone levels and other individual differences.

Skill Acquisition Specialists and the Evolution

of Human Interfaced Systems

The present studies and other studies are building bridges between basic research and applications. These bridges enable specialists in skill acquisition to connect their basic foundations with the process of evolving human-interfaced

systems, including weapons and logistics systems, and with the process of training personnel to expert levels of competence. Wes Regian, Eric Day, a graduate student at Texas A&M, and myself are preparing a theoretical article in which we define the "skill acquisition specialist" as a profession who is educationally well-grounded in the theoretical and empirical traditions of learning and skill acquisition, and who applies that background to training problems. In order to fully capitalize on the new bridges that are emerging from research on skill acquisition, we recommend an expanded role of the skill acquisition specialist in evolving human-interfaced systems. This evolutionary process includes iterative design and development of systems, the training of personnel to use the systems, and certain aspects of the long-term management of these personnel. Based on the new information that skill acquisition specialists can bring to this evolutionary process, we argue for their active participation on a multidisciplinary team of specialist throughout successful evolution of human-interfaced systems, which includes careful task allocation between the system and the human, iterative and ongoing system designing, an effective interface between the system and the human, appropriate operator selection and training, systemic motivation for continued improvements in operator skill, and opportunities for skilled operators to influence system development.

Conclusion

My colleagues and I at the TRAIN Laboratory this summer did

research that faithfully represented the cognitive processes that underlie training in the field, and enabled rigorous analysis of those processes. We also noted that such research bridges a gap between basic research and applications, and we explained how such bridges can expand the role of skill acquisition specialists in the successful evolution of human-interfaced systems.

References

Adams, J.A. (1987). Motor skills. Psychological Bulletin, 101 (1), 41-74.

Glenberg, A.M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. Memory and Cognition, 7, 95-112.

Gouchie, C. & Kimura, D. (1991). The relationship between testosterone levels and cognitive ability patterns. Psychoneuroendocrinology, 16, 323-334.

Green, R.L. (1989). Spacing effects in memory: evidence for a two-process account. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15 (3), 371-377.

Johnston, A.N., Regian, J.W. & Shebilske, W.L. (1994). Observational learning and training of complex skills in laboratory and applied settings. Proceedings of the WEAAP (Western European Association for Aviation Psychology) 21st Conference, Dublin, Ireland.

Kimura, D. (1992). Sex differences in the brain. Scientific American, 267, 119-125.

Schmidt, R.A. & Bjork, R.A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training; Psychological Science, 3 (4), 207-217.

Shebilske, W.L., Regian, Arthur & Jordan, J.A. (1992). A dyadic protocol for training complex skills. Human Factors, 34, 369-374.

Shebilske, W.L. & Regian, J.W. (1992). Video games, training and investigating complex skills. Proceedings of the 36th Annual Meeting of the Human Factors Society, 2, 1216-1300.

Smith S.M. & Rothkopf, E.Z. (1984). Contextual enrichment and distribution of practice in the classroom. Cognition & Instruction, 3, 341-358.

Shute, V.J., Pellegrino, J.W., Hubert, L., & Reynolds, R.W. (1983). The relationship between androgen levels and human spatial abilities. Bulletin of the Psychonomic Society, 21, 465-468.

USING THE SEM-EDXA SYSTEM

AT AL/OEA FOR ANALYSES OF AIRBORNE FIBERS

AND DENTAL WATER LINES

Larry R. Sherman
Department of Chemistry
University of Scranton
Scranton, Pennsylvania 18520-4626

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory
Brooks Air Force Base, San Antonio Texas 78235

August 31, 1994

Larry R. Sherman

USING THE SEM-EDXA SYSTEM
AT AL/OEA FOR ANALYSES OF AIRBORNE FIBERS
AND DENTAL WATER LINES

LARRY R. SHERMAN
Department of Chemistry
University of Scranton
Scranton, Pennsylvania 18510-4626

ABSTRACT

The health hazard posed by airborne asbestos fibers is well documented. Numerous analytical methods for their identification and characterization have been published but analytical procedures for other fibers are not as well defined. This project involved using the Amray 1820 Scanning Electron Microscope equipped with an Electron Dispersed X-ray Analyzer to overcome this deficiency in the asbestos analysis section at the Armstrong Laboratory.

During previous tenures at AL/OEA, the author and his colleagues produced data for a number of standard fibers and established a basic fiber/mineral identification library. From the elemental analyses and cation/anion ratios, they succeeded in characterizing most fibers or minerals, which were submitted to the laboratory.

During the current fellowship the library was expanded and modified, not only for fiber/mineral identification, but also to demonstrate the analytical prowess of the SEM-EDXA equipment for analyzing non-routine samples. Over 100 samples were characterized. A great deal of effort was devoted to determine the biological and chemical buildup in dental water lines in conjunction with the Dental Investigations, AL/AOCD. Initial work with the SEM-EDXA and water analysis section at OEA indicates that the

major water line deposits in the Air Force dental units may be corrosion products (Cu, Ni and Zn crystals). A preliminary communication has been prepared for publication from this investigation.

USING THE SEM-EDXA SYSTEM AT AL/OEA FOR ANALYSES OF AIRBORNE FIBERS AND DENTAL WATER LINES

Larry R. ShermanDepartment of Chemistry
University of Scranton
Scranton, Pennsylvania 18510-4626

INTRODUCTION
Airborne Fiber Study

Although asbestos is currently believed to pose the greatest microscopic fiber health hazard to workers in both the fabricating and users' industries, other inorganic fibers may also cause lung problems (1,2). As the number of samples containing asbestos decrease due to the asbestos abatement programs, the number of samples at the action level from other fibers, such as fiberglass, has increased. Many industrial hygienists have adopted the OSHA permissible exposure limit (PEL) for asbestos, i.e. 0.2 f/cc for an eight hour weighed average and 0.01 f/cc for a 30 minute exposure for all inorganic fibrous material, and recommend monitoring all inorganic fibers until further health hazard data can be established (3). The Air Force uses an action level of 0.01 fibers/cc regardless of the fibrous composition.

The Air Force normal fiber identification procedure uses Phase Contrast Microscopy, NIOSH Method 7400, but this procedure can not distinguish hazardous inorganic fibers from innoxious natural materials. One of the major problems has been separating acicular minerals, fibrous minerals and asbestiform minerals. The NIOSH method is unable to distinguish organic from inorganic fibers, thus many high fiber count samples actually pose no health hazard because the fibers are not lung abatement substances. Furthermore, on the sub-micrometer scale, cleavage

fragments can appear as fibrous material which add further ambiguity. On the other hand, assuming that all fibers are inorganic or even upgrading the fiber count to assume they are asbestos fibers can be an uneconomical evaluation and require much more effort than justified if a complete fiber identification were available. Nevertheless, the cost is too high for analyzing every fiber in an airborne sample; the current project was designed to enhance the use of a scanning electron microscopy-energy dispersed X-ray analysis (SEM-EDXA) method for rapidly analyzing a representative fraction of the fibers in air and bulk samples for evaluating the environment from which the sampling occurred.

During two previous tenures at Brooks AFB, the author had shown that SEM-EDXA is a faster more economic method than a TEM method for identifying asbestos fibers with a greater accuracy for amphiboles and common minerals. He and his coworkers demonstrated that SEM-EDXA can also be used for identifying other airborne materials with a high degree of certainty (4,5). The current tenure was used to demonstrate the use SEM-EDXA as a supplemental analytical procedure for AL/OEA functions and to definitively identify airborne fibers and other microscopic materials.

BACKGROUND

Asbestos contains silicon, magnesium, iron, calcium, and sodium with aluminum replacing some calcium and silicon atoms and manganese replacing some of the iron atoms in the crystal structure. Most natural airborne inorganic materials are calcium, magnesium or aluminum silicates but some samples may contain

sulfur, phosphorus or potassium. In rare situations, they also contain chlorine. Most of these minerals can easily be identified by SEM-EDXA. Man-made fibers, which are now believed to be as dangerous as asbestos (1,2,3,6), are more difficult to identify because of the nearly identical compositions, i.e. $CaSiO_3$ or $Al_2(SiO_3)_3$.

The majority of inorganic minerals in air samples are naturally occurring "dust particles" whose composition depends upon geographic location and environment, i.e. semi-desert and dusty atmosphere contain amosite, a natural asbestos derived from rock weathering. On the other hand, samples obtained in office buildings may contain talc from cosmetics, face powders, body powders and some prescription medicines.

Morphology can be used to distinguish a fibrous from a non-fibrous material but, without chemical analysis, gives little insight into the type of mineral in the samples. Although morphology can distinguish chrysotile from amphiboles, it cannot be used to distinguish individual amphiboles or amphiboles and non-asbestos fibrous materials. These require a complete elemental analysis which the SEM-EDXA instrument is capable of performing.

METHODOLOGY

SEM-EDXA

Standard inorganic fiber samples and minerals, as well as unknown solids, were mounted on carbon coated aluminum SEM studs and overlaid with a Au/Pd coating using the Anatech Ltd. Hummer VI as described in the literature (4,5,7,8). They were analyzed by standard methods and the weight percent used for further

chemical analyses (4,7,8).

All seven asbestos types and over ninety non-asbestos materials were used to establish a library for mineral analyses. Since EDXA unit can only decipher elements with an atomic number 11 or greater, the stoichiometric composition and the library values do not match. SEM-EDXA is a surface analyses procedure and common elemental substitution will also distort the results. Nevertheless, the substitutions are quite reproducible and windows (ranges of composition) have been established from the molar element ratios for quickly identify most materials. Cation/anion ratios were calculated assuming that the electron density is greatest upon the Si, Al Cl, S, and P moieties (anions) and least on Ca, Mg, Na, K, Fe and Mn moieties (cations). With non-asbestos materials, they were not normalized, since the number of atoms per unit cell varies from mineral to mineral. However, the asbestos and other mineral cation/anion ratios have been excellent parameters for positively identifying materials with the same or nearly the same composition. Cation/anion ratios help to correct for natural substitution of one element for another in the crystal structure since the valence state or coordination number does not change.

Unknown samples were searched for suitable fibrous or globular morphology. A characteristic portion of the material was analyzed using the Energy Dispersed X-rays emitted after impinging electrons on the surface of the sample. The ejected photons were analyzed using a Tracor Northern X-ray Analyzer II (TN) and corrected for absorption, interference and atomic num-

ber. Using "canned" programs supplied by TN, the TN computer calculated the raw data as weight percents of the eleven common elements found in natural inorganic materials. A minimum of five fibers/minerals were processed for statistical purposes.

COMPUTER PROGRAMS

Identification of minerals as either fibrous or non-fibrous is extremely tedious with composition tables. To ease the work and use technicians with less chemical expertise, several user friendly computer programs have been prepared for identifying the materials. To be portable, the analytical computer programs were written in C++. They require a 286 or higher computer with DOS 3.3. Since the programs contain thousands of lines, they are not reproduced in this report but are available from either the laboratory focal coordinator or the faculty fellow (4).

ASBT1.CPP PROGRAM

This program processes only asbestos data by comparing the molar element/silicon ratios and cation/anion ratios based upon the library material and real samples. When the program is compiled and mounted in the file manager, the chemist can use the arrow and the numerical keys to identity if a fiber fits the composition of one of the seven asbestos types. The program prepares a laboratory report which is stored as ASBFIBER.TXT and can be accessed from the file manager.

CALCIUM.TXT PROGRAM

Calcium is a critical element in mineral analyses. All standard minerals containing more than 5% calcium, including tremolite, actinolite and ferroactinolite, were put into a single

computer program called CALCIUM.CPP. A range of compositions based on the molar element/calcium ratio and cation/anion ratio was established. Windows were created for each calcium bearing minerals and the program identified most common calcium bearing minerals found in AL/OEA samples.

When complied and mounted in the file manager identification is possible using arrows and the numerical keys. The laboratory report is stored as CALCIUM.TXT and is accessible through the file manager.

HISIL.CPP PROGRAM

HISIL.CPP program is designed for samples containing more than 10% silicon, with or without calcium. Some overlap exists with minerals found within this file and other files. The file contains over 40 minerals including four asbestos, chrysotile, amosite, crocidolite and anthophyllite. All weight percent mineral data are stored in a matrix from which the standard composition tables are prepared. Unknown molar data is manipulated to check against standard molar element/silicon ratios and cation/anion ratios.

The program is accessible with arrow and numerical keys in the file manager; the laboratory report is stored as HISIL.TXT in the file manager.

MISCEL.CPP PROGRAM

The fourth program MISCEL.CPP will process materials with less than 5% calcium, less than 10% silicon or which contain unusual elements, e.g. copper, tin, zinc or titanium. The program is accessible through the file manager and the laboratory report

is stored as MISCEL.TXT. The program is limited but contains minerals that are not common. This program requires further data manipulation to be fully usable. The author will request a mini-grant to finish this work.

BASE.REP PROGRAM

The fifth program was put together for sending the results of a SEM-EDXA analysis to the base submitting a sample. The usable program is written in basic and is less user friendly than the number crunching programs.

A C++ user friendly program is required for completion of the problem. As part of the mini-grant, the author wishes to prepare a report program which is portable and only requires arrow keys for printing. In the mean time, the asbestos function chemists can access the program through the file manager, then press F3 followed by typing B.BAS then press the F2 key. Other information needed to print the report is requested on the screen.

RESULTS.

The SEM-EDXA unit was used by the P.I. for analyzing more than 100 samples (in excess of 500 individual analyses). At a commercial rate of $400/sample this involved approximately $40,000 worth of chemical analyses. The samples consisted of simple air fiber samples with an action limit greater than 0.01 fibers/cc, bulk samples which needed verification, unknown samples received as black smears on toweling and samples to be sized. Some of the air fiber samples contained asbestos, some contained fiberglass and some only contained innoxious materials.

All other samples contained materials which did not pose an environmental health hazard.

Dental Investigations Group

Near the end of the P.I. tenure at Brooks AFB, he was approached by Col. Shannon Mills of the Dental Investigation Section, AL/AOCD, to perform some SEM-EDXA work on the dental water lines at the Brooks Clinic. The X-ray analyses indicated high concentrations of copper, nickel and zinc. The work has become a major project and a preliminary communication has been prepared in regard to the project. The communication is reproduced as follows:

INTRODUCTION

The phenomenon of microbial colonization within the confines of small bore plastic tubing used for dental unit water lines was first reported over 30 years ago (9). Subsequent investigations have revealed much about the nature of the microbial aggregate bio-films responsible for this problem (10,11,12). Factors contributing to the high levels of microorganisms found in dental water lines, which may exceed 10^{+6} colony forming units per mL (CFU/mL), include plastic substrates favorable for colonization by water bacteria, high surface to volume ratios within small bore tubing, and laminar flow characteristics. The most widely used method for controlling dental unit bio-films involves the use of separate water reservoirs which isolate the unit from community water supplies. Periodic disinfection is necessary to control resident bio-films.

Units at two institutional dental clinics have been disin-

fected with a 1% solution of sodium hypochlorite once a week for several years. This procedure has been largely effective in controlling microbial contamination and has not resulted in obvious damage to the dental units. However, sodium hypochlorite apparently causes very low levels of corrosion to the metal portions of dental unit resulting in the formation of mineral deposits on the walls of the water lines. At low magnification, these mineralization can be confused with bio-films; however, the chemical composition is entirely different.

Five sections of dental unit water line were cut from operating dental units in clinical use at two US Air Force dental clinics (usually a section of inlet line, the air-water syringe and three hand pieces 1,2 &3). A 5 mm section of the line was cut longitudinally and mounted on standard aluminum SEM studs. Selected mineralization or bio-films-- which were usually relatively homogeneous in appearance in any single sample were analyzed for Si, Mg, Fe, Ca, Na, Al, S, Ca, P, Cu, Ni, and Zn at 14,000-17,000 magnification for 90 seconds as previous described in the literature (7-8). The photon counts varied from a few hundred to more than 20,000 counts depending upon the type of sample analyzed.

RESULTS AND DISCUSSION

Three distinct morphologies could be seen on the electron micro-graphs at 2600x; they were labeled, hills, mountains and lakes because they had appearances were similar to aerial photographs.

The lakes were dental water line tubing. They contained

trace copper (ppm) or in rare cases iron (never both) which probably originated from the water inlet pipes connected to the dental unit. If the unit was free of bio-film, they also contained trace nickel and zinc.

Results of the analyses of the hills varied depending upon the water line treatment. Water lines connected to community water supplies contained deposits morphologically and chemically consistent with microbial bio-film with traces of calcium, silicon and copper. Bacteriological analysis confirmed the presence of high levels of bacteria in these samples. Water lines samples taken from three units at one of the two clinics, which had been undergoing routine sodium hypochlorite treatment for over four years, did not exhibit deposits consistent with bio-film contamination. Bacteriological samples confirmed the absence of significant levels of bacteria in these units (0-2 colonies/mL). Representative deposits in these water lines contained high concentrations of copper, nickel and zinc; the later two metals constituted about 80% of the metal in the coatings. AAS and ICP analysis of the water from the various lines yielded confirmatory data. Units from the second clinic, that were also undergoing intermittent disinfection, contained some bio-film (an observation confirmed by bacteriological analysis). Hill deposits in these latter lines contained relatively high concentrations of copper and zinc with small amounts of nickel in the deposits.

The mountains, which were analyzed as mineral deposits, were found only in samples free of bio-film and appear to be mixed

crystals of copper, nickel and zinc with nickel being the predominate metal in the deposits. These crystals are most likely oxides. The nickel and zinc conceivably came from hypochlorite corrosion of the dental unit. Chlorine content in any sample depended upon the time elapsed since the last disinfection procedure on the unit.

Table 1 lists the SEM-EDXA data for a number of dental water lines. The presence or absence of a bio-film was determined by SEM, light microscopy and microbiology of the water from the lines. The metal analysis is the average of 3-5 analyses on different particles in different portions of the tubing. H,M or L refers to the number of photons detected by the detector in 90 seconds. Other elements, which are found in most dirt samples, are listed when they yielded a moderate number of photon counts. The percent chlorine is given when it constitute more than 2% of the particle analyzed. When bio-films were present the analysis corresponds to hills; when no bio-films were present the analysis corresponds to mountains. When the inlet line area are removed from the statistical analysis of the data, there is a direct correlation between the amount of nickel present in the analysis and the amount of bio-film on the water lines (p=0). The inlet water lines exhibited erratic results because some had been cleaned and others had not. The current data available to the investigators was incomplete to a[[raise inlet water chemical and biological deposits.

FUTURE WORK

Although the fiber/mineral analysis programs have reached the level that they can be used in routine analyses of samples which have "action level" concentration of fibers, some work is still needed to edit the computer programs so that a technician can use them.

A great deal of work is required to understand the chemistry and microbiology of dental water lines. The work could involve hundreds of analyses and will likely be continued on an intermittent basis for the next two to three years.

REFERENCES

1. N.A. Esmen, Y.Y. Hammad "Recent Studies of the environment in ceramic Fiber production. In: Biological Effects of Man-made Mineral Fibers, Vol. 1, pp 222-231. WHO/IARC Publication Lyon France (1984).

2. K.H. Kilburn, D. Powers & R.H. Warshaw, "Pulmonary Effects of Exposure to Fine Fiberglass: Irregular Opacities and Small Airways Obstruction," Br. J. Ind. Med. **49**:714-720 (1992).

3. R.T. Cheng, H.J. McDermott, G.M. Gia, T.L.G. Cover & M.M. Dude "Exposures to Refractory Ceramic Fibers in Refineries and Chemical Plants" Appl. Occup. Environ. Hyg. 7, 361-7 (1992).

4. L.R. Sherman "Final Report Summer Fellowship," US Air Force of Office of scientific Research (1993).

5. G. Monka "Final Report Summer Fellowship," Using the Scanning Electron Microscopy for Fiber Analysis, US Air Force Office of Scientific Research (1993).

6. J. Zussman, "Crystal Structures of Amphiboles and Serpentine Minerals," P.D. Lafleur & K.J. Heinrich ed., Nat. Bur. of Stand. Washington, 20234 (1978).

7. L.R. Sherman, K.T. Roberson & T.C. Thomas "Qualitative Analysis of Asbestos Fibers in Air, Water and Bulk Samples by SEM-EDXA" J.PA Acad. of Sci. **63**, 28-33 (1989).

8. K.T. Roberson, T.C. Thomas & L.R. Sherman, "Comparison of Asbestos Air Samples by SEM-EDXA and TEM-EDXA" Ann. Occup. Hyg. **36**, 265-9 (1992).

9. Blake, G.C. "The incidence and control of bacterial infection in dental spray reservoirs." Br. Dent. J. 1963 **115**:413-416.

10. Kelstrup, J., Funder-Nielsen, T.D., and Theilade, J. "Microbial aggregate contamination of water lines in dental equipment and its control." Acta Pathol. Microbiol. Scand. [B] 1977 **85**:177-183.

11. Mills, S.E., Lauderdale, P.W., and Mayhew, R.B. Reduction of microbial contamination in dental units with providone iodine 10%." J. Am. Dent. Assoc. 1986 **113**:280-284

12. Williams, J.F., Johnston, A.M., Johnson, B., Huntington, M.K., and Mackenzie, C.D. "Microbial contamination of dental unit waterlines: Prevalence, intensity, and microbial characteristics." J. Am. Dent. Assoc. 1993 **124**:59-65

TABLE 1
Analysis of Dental Water Lines

#	Type of Tubing	Bio-film	Cu	Ni	Zn	Other elements,
4516	I	Y	H(86)			Ca/Si
4529	I	Y	L(41)	L(21)		Al
NONE	I	Y	M(75)	L(6)		Si/Cr
5737	I	Y?	M(14)	M(16)	M(37)	Cl
4524	I	N	M(78)			Na/Al
5263	I	N	H(40)	L(2)		Cl(38)/Si/Fe
5266	I	N	H(46)	M(3)		Cl(48)
5268	I	N	M(8)	L(0.5)	M(11)	Cl(78)
4517	A	Y	L(55)	L(21)		
4521	A	Y	H(92)			Al
4528	A	Y	H(9)	H(30)	H(60)	
5771	FOP	Y	M(18)	H(36)	H(39)	Cl
5738	A	N	M(19)	L(4)	M(73)	
5264	A	N	M(50)	L(5)		Fe/Cl(22)
5265	TS	N	L(3)	L(2)		Al/Si
5267	A	N	H(47)	M(5)	M(6)	Cl/Si
5269	A	N	H(64)	M(4)		Cl(11)/Si
5270	FOP	N	M(7)			Cl(30)/Si/Al
4515	H1	Y	H(32)	H(27)	H(30)	
4518	H3	Y	M(17)	H(41)	M(28)	Si/Al
4519	H3	Y(dried)	M(8)	H(43)	M(8)	Na
4525	H1	Y	M(3)	H(45)	H(12)	Na
4527	H2	Y	L(7)	H(44)	M(47)	Al
4520	H3	N	L(5)	M(44)	M(26)	Na
4522	H2	N	M(2)	H(53)	H(16)	Na
4523	H1	N	H(38)	M(17)	M(30)	Na/Cl(7)
5735	H1	N	M(28)	M(9)	H(54)	Si
5736	H2	N	M(7)	M(8)	M(43)	Si
5994	H1	N	M(8)	L(1.5)	H(85)	
5995	H2	N	M(70)	L(10)		Al
5996	H3	N	M(60)	L(3)	L(16)	
5997	H1	N	L(2)	L(1)	M(28)	Cl(40)/Na(20)

The value in parenthesis is the normalized % of that element in the total package of elements analyzed.

I = Inlet line
A = Air Water Syringe
H = Hand-piece

L = Low Photon Counts (<100)
M = Moderate counts (100-1000)
H = High Counts (>1000)

FOP = Fiber Optical Piece
TS + Three way syringe

words 3654

REGIONAL ARTERIAL COMPLIANCE AND RESISTANCE CHANGES

FOR TRANSIENT +Gz PROFILES

Richard D. Swope

Professor

Department of Engineering Science

Trinity University

715 Stadium Drive

San Antonio, TX 78212-7200

and

Daniel L. Ewert

Assistant Professor

Department of Electrical Engineering

North Dakota State University

Fargo, ND 58105

ABSTRACT

The primary aim of this research is to determine regional variations in peripheral resistance, blood volume and arterial compliance caused by transient +Gz loads. A model previously used to analyze systemic arterial compliance and total peripheral resistance is extended to allow similar calculations for the head, lungs and body as well as shifts in blood volume between these regions. Gravitational loss of consciousness (G-LOC) is a direct result of a prolonged blood volume shift from the head to the body and the new model allows a study of the relationship between this shift and regional changes in resistance and compliance on a beat to beat basis. Practical surgical limitations require the development of a new transducer for measuring pressure and flow in the pulmonary artery and the aorta before the method can be implemented. Preliminary work with a modified transit time ultrasonic transducer shows promise as a solution to this problem.

Background

The vulnerability of cardiovascular function to the increased gravitational (+Gz) loads seen in military aircraft has spawned a large body of applied research over the last 50 years. In recent years the need to wear chemical and biological protective gear has increased this vulnerability because the aircrew often become dehydrated as a result of the added thermal load. Under normal (hydrated) circumstances +Gz loads lead to a pooling of the blood volume in the lower body and legs and a consequent decreased stroke volume of the heart. This limits the ability of the baroreflexive mechanisms to restore arterial cerebral perfusion pressure and if the blood supply to the brain is reduced for a long enough time loss of consciousness (G-LOC) results. With dehydrated crew members plasma volume is reduced and it is expected that the drop in stroke volume will occur at lower +Gz loads. In fact, given enough dehydration, one would expect problems even at the +Gz loads found in some helicopter maneuvers, (3 to 4 G).

Total peripheral resistance (TPR) and systemic arterial compliance (SAC) are two parameters which may be used to help understand and ultimately provide solutions to, the problems of +Gz loads. TPR, a measure of the hydraulic resistance or vascular load of the left ventricle during ejection, is a significant player in determining aortic blood pressure. SAC, a measure of the elastance of the arterial system is also significant; together they determine the

systemic impedance load for the left heart and the pressure/flow characteristics in the aorta.

Researchers historically have had difficulty translating physiological signals into useful descriptions of short-term systemic pressure regulation, especially under non-stationary conditions. Underlying this has been an inability to determine TPR and SAC during non-steady-state periods of +Gz exposure. Standard methods that use mean pressure and flow (1) are inappropriate under transient conditions because of the varying amounts of blood stored in the arterial compliance (2). As a result non-steady-state TPR can only be correctly derived when SAC is taken into account.

We have developed a method (3) which provides beat to beat values of TPR and SAC under transient conditions similar to those found in military aircraft. Modifying the model to include separate vascular beds for the head, lungs, and remainder of the CV system will allow a determination of regional blood volume shifts during transient +Gz loads and a better understanding of the effects of hydration state.

Proposed Model for Dehydration Studies

Our previously developed model for computing TPR and SAC is a two element Windkessel consisting of a lumped systemic arterial compliance (C_{ao}) and a lumped resistance ($R_{arterial}$) as shown in Figure 1.

41-4
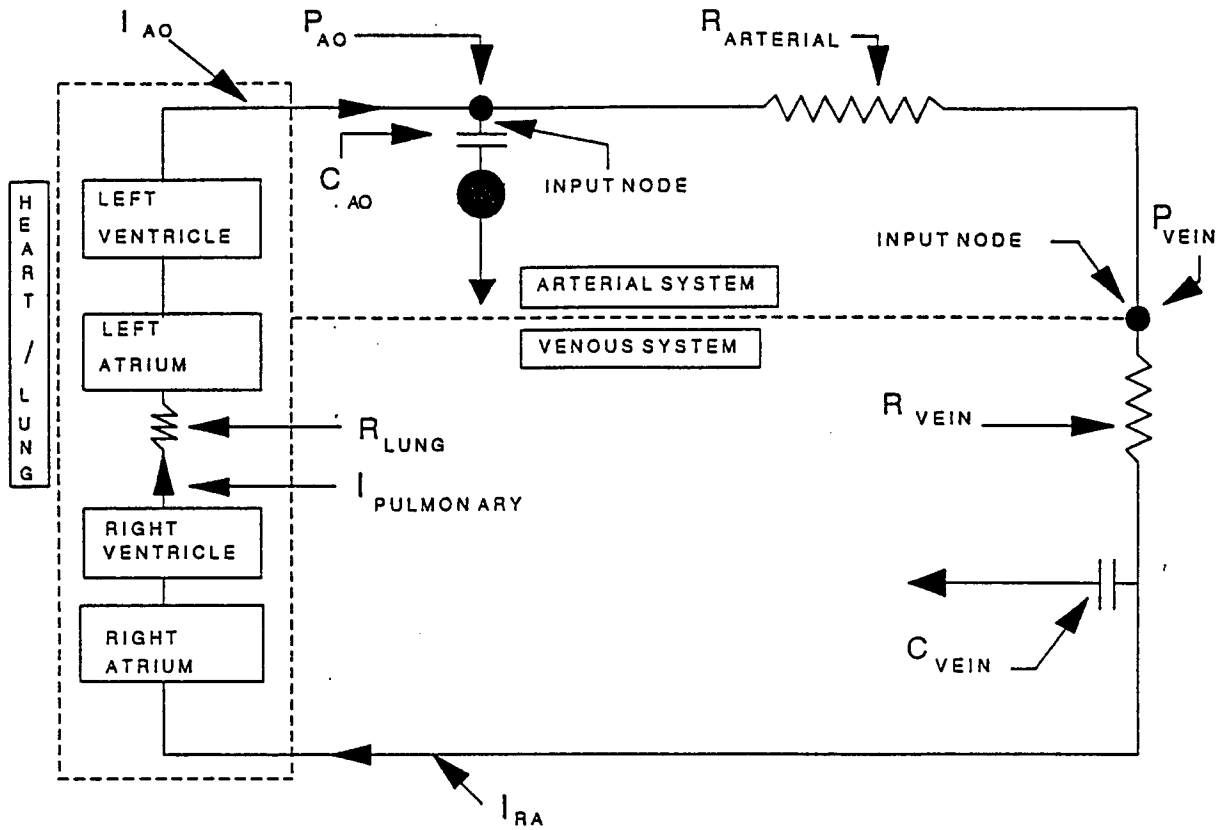
Figure 1

By applying Kirchhoff's current law to the input node and translating the result to cardiovascular terms we get

$$I_{ao} = C_{ao}\frac{d(P_{ao} - P_{pleural})}{dt} + (P_{ao} - P_{ra})/R_{arterial} \qquad (1)$$

where,

$$C_{ao}\frac{d(P_{ao} - P_{pleural})}{dt}$$

represents that portion of the aortic flow which goes into charging the aortic capacitance (P_{ao} = aortic root pressure, $P_{pleural}$ = extravascular thoracic pressure) and ($P_{ao} - P_{ra}$)/$R_{arterial}$ is the flow

through the resistor. (P_{ra} = right atrial pressure, is taken to be equal to venous pressure after correcting for hydrostatic offset). The computational details of the procedure for calculating TPR and SAC are given in reference 3. It is noted here that the evaluation of these parameters requires continuous recordings of aortic root pressure, aortic root flow and right atrial pressure.

The proposed model uses the same basic elements as above for studying the effects of hydration states but distributes them into three regions as shown in Figure 2.
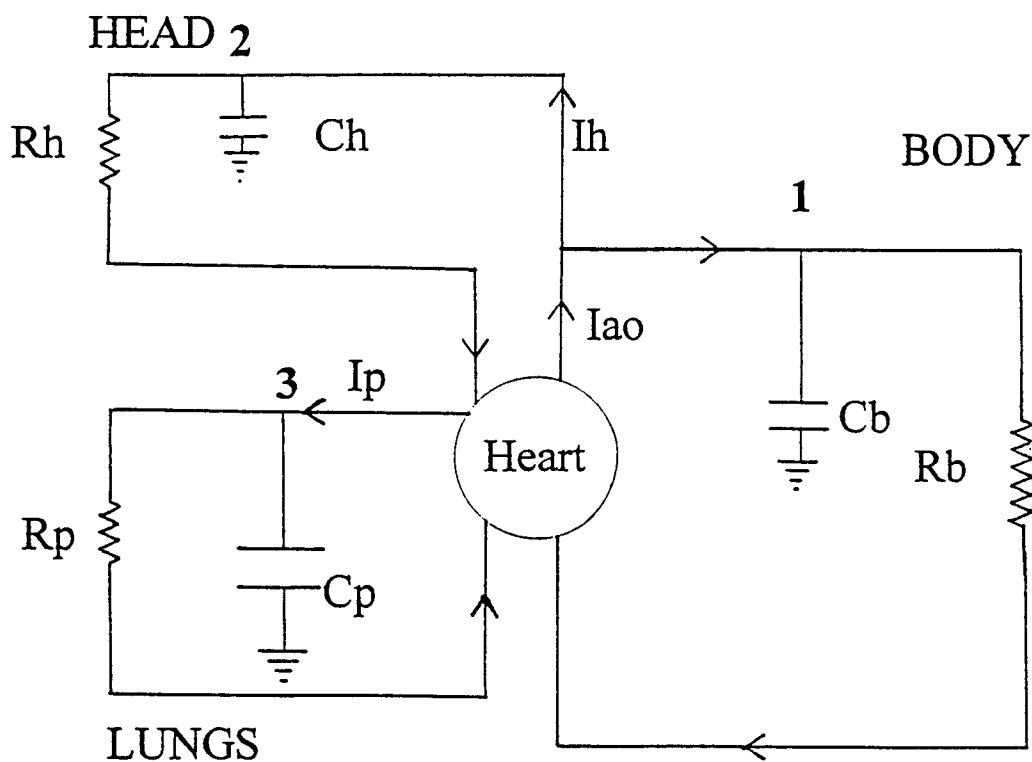
Figure 2

Thus, the head, lungs and body are each represented by a two element Windkessel. This is the simplest model suitable for studying fluid volume shifts and changes in vascular bed properties. It requires the continuous measurement of aortic root flow (I_{ao}) as well as flow to the head (I_h) and Pulmonary artery flow (I_p). Required pressure measurements include: aortic root pressure, Pulmonary Artery pressure, plural pressure and right and left atrial pressures.

Applying Kirchhoffs current law to nodes 1, 2 and 3 of Figure 2 and translating the results into cardiovascular terms we get.

for node 1

$$I_{ao} - I_h - C_b \frac{d(P_{ao} - P_{pleural})}{dt} + (P_{ao} - P_{ra})/R_b \qquad (2)$$

for node 2

$$I_h = C_h \frac{d(P_{ao} - P_{cran})}{dt} + (P_{ao} - P_{ra})/R_h \qquad (3)$$

for node 3

$$I_p = C_p \frac{d(P_p - P_{pleural})}{dt} + (P_p - P_{la})/R_p \qquad (4)$$

Where P_{cran} = cranial pressure and P_{la} = left atrial pressure and C_b C_h, and C_p are compliances for the arterial components of the body, head and lungs respectively. R_b, R_h and R_p are the corresponding resistive elements. With corrections for hydrostatic offset and the

measurements mentioned above, the procedure used in our earlier two element model may now be applied to each model equation to generate beat to beat values for C_b, C_h, C_p, R_b, R_h and R_p. Furthermore, by integrating the flow rates I_{ao}, I_h and I_p and subtracting the results we can quantify blood volume shifts (between the three regions of the cardiovascular system) over the course of transient +Gz load variations.

The added information available with this model comes with a price. In our first model only two pressures and one flow rate needs to be measured. In our proposed model it is necessary to continuously measure 6 pressures and 3 flows. It might be possible to reduce the number of measured pressures by making some assumptions about changes in pleural and cranial pressures which are relatively small compared with aortic and pulmonary artery pressures. Even if these assumptions can be made, four pressures and three flows will be needed and thus surgical limitations must be considered. Personal communications with Dr. John Fanton of the Armstrong Laboratory indicate that the only practical problems to be overcome are associated the measurement of pulmonary artery flow and pressure and flow and pressure in the aorta at the same time. The problem is basically one of "real estate". There is not enough room to place all of the needed transducers. Thus, there is a need for a new transducer which is smaller than the ones we have used in the past. The development of such a transducer is the goal of an AFOHSR proposal from Dr. Dan Ewert (4). Other goals of this proposal include modified surgical techniques intended to dramatically

reduce recovery time and costs.

We have completed in-vito testing of a relatively new Triton active-redirectional transit time (ART^2) flow transducer to determine if it is suitable as a basic foundation for further development. Our intention is to incorporate a pressure transducer in the housing of the flow transducer to allow simultaneous extravascular measurement of pressure and flow.

Two adult rhesus monkeys were instrumented with Triton ART^2 probes, and flow readings were recorded (and calibrated against Thermal Dilution (TDL) measurements) for 5 days over a 3 month period. On each day a total of fifteen recordings took place: 5 baseline TDLs, 5 TDLs after administering nitropruside and 5 TDLs after administering phenelephrine. The drugs were used to produce a wide cardiac output range.
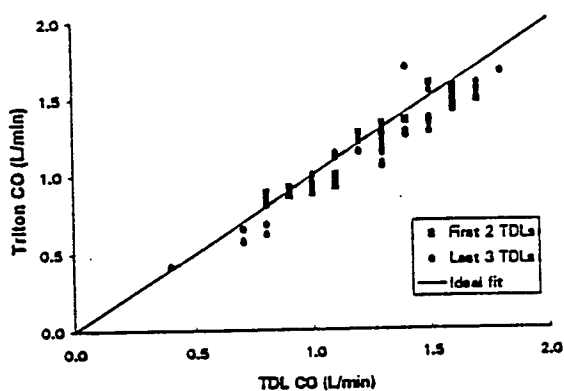
Figure 3
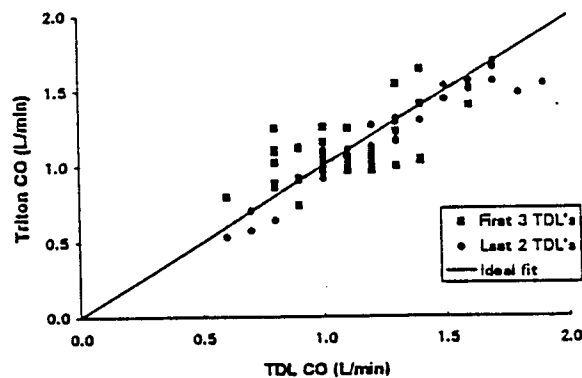
Figure 4

The results of the study are shown in Figures 3 and 4. Figure 3 is for one test animal and 4 is for the other. In each figure a straight line with a slope of 1 is drawn through the origin. The data generally fit the line fairly well. It should be noted that the accuracy of the TDL method is about ±15% which could account for some of the scatter shown. Bench tests of the same flow probes using a bucket and stop watch method for calibration showed them to be linear and accurate to ±10% over the flow range shown in the figures.

Wet Lab Verification of Model

TPR and SAC values computed with our first model are comparable to those found for the steady state condition computed by other methods. This agreement increases our confidence in the validity of our approach but since there is no way to independently measure these parameters in-vivo we cannot be 100% sure of our results. One way of approaching a 100% confidence level is to compare predicted values of TPR and SAC with known (measured) values in an in-vitro setting.

During this research period we have continued the development of a "Wet Lab" (5) which can be used as a test bed for evaluating transducers under controllable and realistic conditions. The facility is now developed to the point where it can be used to validate our transient 2-element windkessel model method. We can program a servo controlled pump system to produce realistic aortic

pressure waveforms. Figure 5 shows measured left ventricular and aortic pressures as well as the pump piston displacement (LVDT) as functions of time.
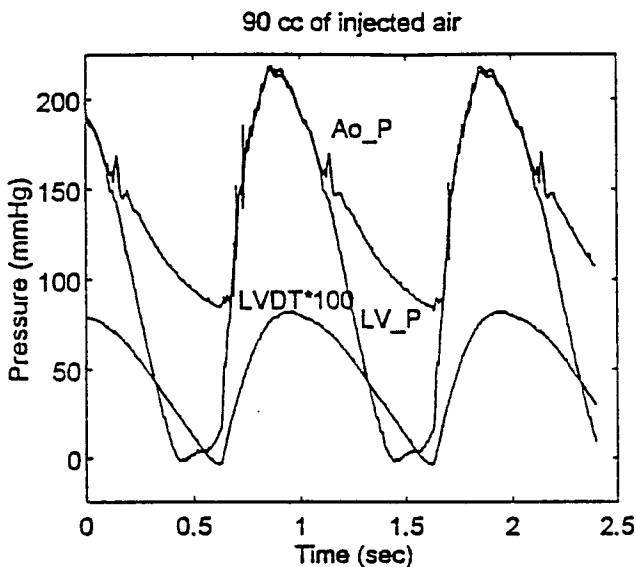
Figure 5
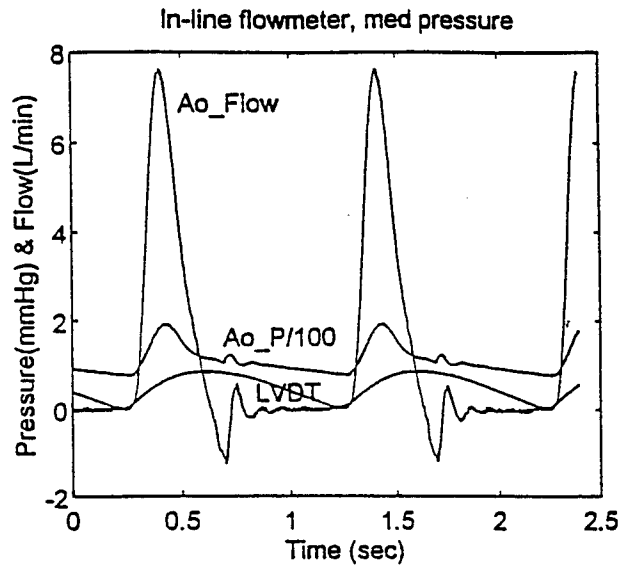
Figure 6

While the pressure peaks are a bit high, the waveforms are realistic and we have learned that the peaks are influenced by the amount of air trapped in the pump system "left ventricle". Future efforts include introducing a larger air cushion to reduce the peak pressure to the 130 to 140 mm Hg range. Figure 6 shows aortic flow superimposed on the pressure waveform and the LVDT trace. The flow is very realistic. Our pump facility has many adjustable parameters and we have not yet found a combination which simultaneously produces realistic pressure and flow waveforms simultaneously. We do however, understand the influences and should be able to do so

with little additional work.

To complete the in-vitro validation we need to construct an experimental model of the two element Windkessel using elastic tubing for the capacitance and an adjustable needle/plug valve for the resistance. For realistic simulation the tubing should have a compliance in the range of 0.5 to 2.0 cc/mmHg and the resistance range should be adjustable from about 1000 to 5000 dyne sec/cm^5. Figures 7 and 8 show the calibration of candidate tubing and resistance valves. Three lengths of latex rubber tubing were tested (1,2 and 3 ft). For pressures in the range from 50 to 150 mmHg the compliance is a linear function of pressure and length. One would expect compliance to be a linear function of length for any pressure range and our data seem to bear this out for the limited range of the tests. A 1 to 2 foot length of this tubing should be appropriate as the human capacitance equivalent for our two element Windkessel model of SAC.
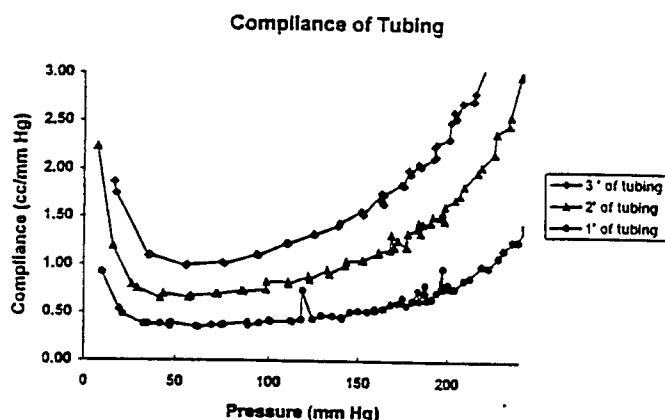
Figure 7
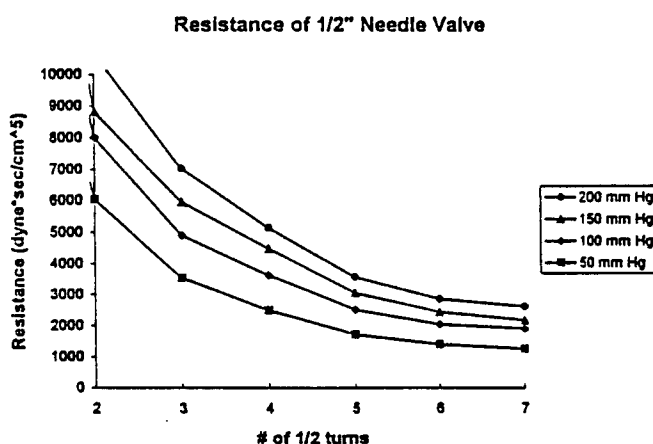
Figure 8

Figure 8 shows that a 1/2" needle valve opened two to three turns gives an appropriate resistance for a range of pressure drops from 50 to 200 mm Hg. Even though the resistance is a function of pressure drop it is expected to be fairly constant over the course of a beat because of the windkessel effect. It is expected that the valve will see an input pressure approximately equal to the mean "aortic pressure".

A student (Jeremy Schaub) from Trinity University will continue to work on this project part time during the 94/95 academic year. He will construct a flow loop model of the arterial system shown in Figure 1 (i.e. the 2 element windkessel) and measure the input node pressure and flow variations for pump produced aortic pressures and flows. These data will then be used in conjunction with our beat to beat method (2) for computing SAC and TPR. A comparison of the computed and measured values will determine the "goodness" of our method.

If our method is proven valid (we think it will be) a flow loop equivalent of Figure 2 will be constructed on a tilt table and used to study regional fluid volume shifts and regional changes in SAC and TPR.

Conclusions

The two element windkessel is an adequate but simple model for describing resistance and compliance changes associated with

transient +Gz loads. By representing the head, lungs and remaining body as a combination of three two-element windkessels regional fluid volume shifts under transient +Gz loads may be determined.

The Triton ART^2 flow probe is suitable for our needs to measure arterial blood flow. It also is suitable for adaptation to measure pressure but additional development is needed.

References

1.   Nichols, W.W., O'Rourke, M.F., McDonald's Blood Flow in Arteries, Lea and Febiger, 3rd Edition, 1990.

2.   Toorop, G.P., Westerhof, N., and Elzinga, G., Beat-to-Beat Estimation of Peripheral Resistance and Arterial Compliance During Pressure Transients, Am. J. Physiol., 1987, 252: H1275-H1283.

3.   Self, D.A., Ewert, D.L., Swope, R.D., Crisman, R.P. and Latham, R.D., Beat-to-Beat Determination of Peripheral Resistance and Arterial Compliance During +Gz Centrifugation, Aviation, Space and Environmental Medicine, May 1994.

4.   Development of an Extravascular Pressure/Flow Transducer, AFOSR Summer Research Extension Program Proposal; submitted by Dr. Dan Ewert, August, 1994.

5. Development of an Enhanced Hydraulic Cardiovascular Model/Test
    Apparatus for In-Vitro Simulations in Altered-G Environments.
    R.D. Swope, Final Report, USAF/RDL Summer Faculty Research
    Program, September, 1991.

REPRESENTING AND TEACHING A DISCRETE MACHINE: AN INSTRUCTIONAL
DESIGN PROBLEM IN PROCEDURAL LEARNING

Steven D. Tripp
Assistant Professor
Department of Curriculum and Instruction

University of Kansas
Bailey Hall
Lawrence, KS 66045

REPRESENTING AND TEACHING A DISCRETE MACHINE: AN INSTRUCTIONAL DESIGN PROBLEM IN PROCEDURAL LEARNING

Steven D. Tripp
Assistant Professor
Department of Curriculum and Instruction
University of Kansas

Abstract

My Casio AltiDepth Watch is emblematic of a class of modern machines that we are faced with mastering. It has 51 functions controlled by four buttons. The internal operations of the watch and the pattern of button-presses are obscure to the user. Learning to operate such machines is notoriously difficult. Standard instructional design (ID) prescriptions for teaching procedures are not helpful. ID typically conceives of procedures as a single string of salient events, perhaps with decision points. The Casio watch is not a single procedure, but a set of parallel procedures, each of which is minimally different from the others. Although there is visual and auditory feedback from the display, it is not sufficient to support operation without some degree of mastery of the command language. For this project a computer-based simulator was developed and a task-action grammar (TAG) was written to represent the command interface. The simulator and TAG will be used to test hypotheses about different approaches to designing instruction for discrete machines; i.e., machines operated by buttons or keys that assume a set of discrete states.

REPRESENTING AND TEACHING A DISCRETE MACHINE: AN INSTRUCTIONAL DESIGN PROBLEM IN PROCEDURAL LEARNING

Steven D. Tripp

Introduction

My Casio AltiDepth Watch is emblematic of a class of modern information machines that we must learn to operate. It has at least 51 functions controlled by four push-buttons. Among its functions are an altimeter, a barometer, a depth meter, an alarm, a countdown watch, and a stopwatch. The operation of the watch is through simple presses on one of four buttons labeled Mode, Adjust, Split-Reset, and Start-Stop. The user receives feedback through an LCD and a beeper that provide some indication of the results of user action but do not indicate what should be done next. Different operations are produced by sets of actions that are only minimally different and sometimes confusingly inconsistent. The mastery of machines such as this is notoriously difficult, witness the number of jokes about programming VCRs. Brown (1986) noted this learnability problem and catalogued some reasons. Browns' first reason was system opacity. Unlike most mechanical devices, information machines are opaque—their structure gives no sense of their function. Mechanical devices, on the other hand, are often operated with physical actions that, through their speed, direction, resistance, noise, etc., give implicit cues as to what is happening internally. Compare a lawn mower with a computer screen. The button-presses that control the computer give no sense of what is happening inside. Likewise the screen changes, beeps and whirring sounds of the computer are uninterpretable without considerable understanding of what a computer is.

Brown's second reason for the difficult learnability of information machines is that they exhibit complexity due to multiple processes. Unlike traditional mechanical devices, information machines can execute several processes at once. My Casio watch can keep clock time, countdown from 15 minutes, time an event, and measure the barometric pressure at once. The user is presented with the problem of understanding what state the watch is in when it may be in several states at once. Some of these states may interact with others. If the clock timer is in 24 hour mode then the alarm mode will be also. If the altimeter is set to meters, then the depth meter will be also. The user manual does not explain all of this, nor does the LCD indicate these interactions.

The learnability problem is further complicated by functional complexity. Traditional mechanical devices typically have a very limited range of functions. Barometers measure air pressure. Can openers open cans. Information machines have multiple layers of functionality,

often only weakly related. Measuring the altitude and timing a runner may have some theoretical relationship, but asking the same machine to do both is conceptually confusing. The user must assimilate the notion that virtually indistinguishable actions can produce radically different results.

As Brown points out, the user is hard put to generate any useful metaphors to explain such machines. No combination of wheels, pulleys, and levers will adequately represent information machines. Since information machines are interactive machines, the user must not only issue commands but must also interpret the results of that action. Although many computer interfaces are now extremely well-developed, my Casio watch is highly limited in the information it can display. For that reason the user is constantly forced to interpret ambiguous displays and act based upon inadequate information.

To summarize, my Casio watch is representative of a class of machines that are difficult to learn. Thus it provides a convenient challenge to instructional designers—how can we best design instruction for such machines?

The problem

Given that modern information machines are difficult to learn, how should we design instruction to help users master such machines? The standard texts are not very helpful. Gagné, Briggs, and Wager (1992) consider this type of learning to be a kind of intellectual skill, rule learning. Learning a rule requires that students' behavior exhibit a "regularity" over a variety of situations. The internal condition of rule-learning is that the student be able to recall the component concepts of the rule. The external conditions are that the instructor verbally indicate the order of the component concepts. The amount of verbal guidance may be more or less lengthy. The instructor may reduce the amount of verbal instruction and have the students discover the correct order of the component concepts. Rules are considered to be isolated units of study.

Gustafson and Tillman (1991) mention that there are at least eight ways to sequence instruction: chronological, order of performance, known to unknown, taxonomical, simple to complex, easy to difficult, interest to learners, and availability of resources. None of these is linked to a particular category of learning. Within-lesson instruction should follow a beginning-middle-end pattern. If you have chosen an expository strategy, they recommend that content be taught in this order: facts, concepts, rules, and problem-solving. Rule learning is not considered in any detail.

Okey (1991), basing his recommendations upon Gagné's nine events of instruction, states that for rule learning the student should be given an example of the rule, cues to the proper sequencing and an opportunity to practice (with feedback) new instances of the rule. Again rules are considered to be isolated units of study.

Smith and Ragan (1993) classify the type of learning involved in mastering an information machine under the category of procedural rule. Procedural rules are procedures. Procedures may be simple or complex. Simple procedures are those with a small number of steps and without branches and they are usually called linear or serial. Complex procedures are those with many steps or with branches. They are called branching or parallel. According to Smith and Ragan simple procedures are taught step-by-step or in reverse order, whereas complex procedures are simplified by teaching first the simplest or most common path, the major branches, or a simplified initial case. Smith and Ragan note that a didactic approach is preferred over discovery. Again rules are considered to be isolated units of instruction.

Leshin, Pollock, and Reigeluth (1992) conceive of a procedure as "...an ordered set of actions to achieve a goal. There is often more than one procedure for achieving a given goal" (p. 170). Their general strategy for instruction is presentation, practice, and feedback. They also recommend the following instructional tactics. Present divergent cases to encourage generalization to new situations. Give practice that represents the full range of cases the learner will encounter after instruction. The sequencing of examples should proceed from easy (familiar, concrete)-to-difficult (unfamiliar, abstract) and practice should follow a similar sequence. Alternatively, they recommend that easy examples and practice be presented together, before difficult examples and practice. For instructional enrichment they prescribe attention focusing by pointing out key aspects of examples or practice. As for representation form, they recommend that examples and practice be as close to their post-instructional form as possible. Simpler representations may be better at early stages. If the procedure requires automaticity then: 1. Teach the last parts of the procedure first, 2. present examples and generality, 3. practice with corrective feedback until accuracy criterion has been reached, 4. present speeded drill with R/W feedback, 5. give integrated practice with simultaneous task. Start with mild overload and proceed to high overload. Leshin, Pollock and Reigeluth also consider rules to be isolated units of instruction.

Jonassen, Hannum, and Tessmer's (1989) handbook of task analysis procedures presents several chapters dealing with the analysis and representation of procedures. Chapter five deals with procedural analysis which is used for tasks which consist of a series of overt steps. The task is represented by a flowchart. Chapter six presents an information-processing analysis

appropriate to a task that is a series of steps. This technique is recommended for tasks with covert mental steps and is based upon an analysis of the way an actual competent performer completes the task. Chapter seven presents path analysis. This technique is used for tasks that have a wide variety of paths. A hierarchy of paths is constructed by noting complexity and redundancy. An instructional sequence can be derived from a complexity hierarchy. Of these three techniques only path analysis considers rules to be more than isolated units. Path analysis can detect hierarchical relationships, but it cannot detect horizontal relationships of the type found with information machines such as the Casio watch.

As can be seen from the above discussion, current instructional design thinking does not address the problem of devices that have multiple parallel functions and are controlled by a command language that exhibits few mnemonic qualities. All the above writers consider rule learning to be essentially the learning of a discrete series of steps associated with a single task. None of those writers considers the problem of learning a parallel set of tasks that are minimally different and may be represented internally in a form more abstract than as a series of steps. Only path analysis approaches this problem, but offers no way of representing formally the relationship between similar tasks.

Clearly, there is a need for better instructional strategy prescriptions for the design of lessons dealing with information machines. Two things are needed for the development of such strategies. First, we need some information machines that allow us to test hypotheses about instructional strategies. We could use actual machines, but since such machines can easily be simulated by computer, it is more practical to write a simulator and use it with potential students. The simulator has the added benefit that it can record student behavior in real time and in a form that can be analyzed later. The second thing we need is a theoretical framework that affords the construction of testable hypotheses. Task-action grammars (TAG) have been demonstrated in the past (Payne and Green, 1989) to make superior predictions about the learnability of various "task languages" or computer interfaces. The present project consisted of the development of those two things: the Casio simulator and a TAG that captures the underlying structure of the Casio control language.

Methodology

This project consisted of two parts: (a) the development of a Casio AltiDepth watch simulator, and (b) a TAG that represents the competence of a skilled user of such a watch.

1. The Simulator

The Casio simulator was written on a Macintosh computer in the HyperCard 2.2 environment. The HyperCard 2.2 environment afforded several advantages. First, the author was fluent in the Hypertalk programming language. Second, HyperCard affords the rapid prototyping of computer interfaces. Third, HyperCard 2.2 supports color, unlike previous versions. Fourth, HyperCard 2.2 allows the creation of stand-alone applications. Such applications can be distributed and used on any Macintosh computer without the HyperCard program.

A copy of the simulator appears in Figure 1. The simulation is not a perfect replica of the watch, but is designed to be "informationally" equivalent. The analog watch, which is not an essential part of the information system, is represented to enhance the realism of the simulation. Obviously, the altimeter, barometer and depth meter modes of the watch could not be implemented as functioning devices on the computer. However, the reference levels and alarm functions of these modes are implemented and are adjusted just as with the real devices.
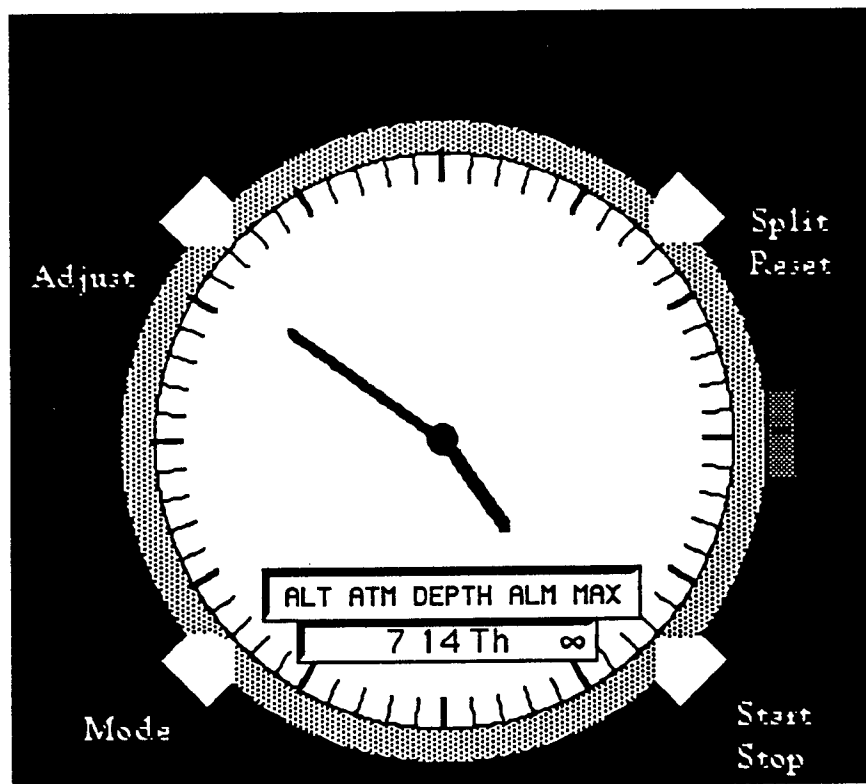
Figure 1. The Casio Simulator

The labels (Mode, Adjust, etc.) that appear beside the simulated watch also appear on the actual watch, but on the face of the watch they are inside and near the four yellow buttons.

Some other changes were made for cosmetic or programming convenience. For example, the symbols that indicate either the alarm or the hourly chime were substituted for simply because no comparable symbol was available in the standard font.

2. The TAG

A TAG (specifically a D-TAG or Display-TAG) was written for the Casio watch and is reproduced in the appendix. The complexity of this notation is necessitated by the complexity of the device. The problem of task representation is not apparent until a device is sufficiently complex. Interactive devices like simple radios or telephones are easy enough to learn that task analysis is trivial. Such is obviously not the case with complex systems. The problem of extending an analysis that succeeds at a relatively simple level to a higher level is exemplified by Polson (1987). Polson described a Cognitive Complexity Theory (developed with David Kieras) that assumed tasks could be decomposed into unit tasks. The theory predicts that training time is a function of the number of new rules that a user must acquire to perform a task. Old rules are assumed to transfer immediately and transfer is based on a common elements "Thorndike-and-Woodworth-type" theory. The theory has been tested on certain simple computer interfaces and has produced a good fit although it is too sensitive to the initial task. The crucial element in this theoretical problem is the definition of rule. Rules are defined at the literal keystroke level. Polson admits this leads to problems and notes that the theory cannot deal with underlying representations of groups of rules that bear a familial resemblance. The general solution to this problem is a grammar to represent rules at the underlying level. TAG is one approach to capturing underlying similarities at an abstract level.

Action Grammars and the Formal Representation of Interfaces.

Action grammars are a kind of task analysis. They are typically applied to computer software interfaces because the complexity of the interface dialogue makes specification a non-trivial task. However, action grammars could be applied to any type of complex task that involves discrete actions and predictable results.

Reisner (1977) introduced the term action language to describe the command system of interactive devices. Reisner was interested in the formal description of such languages. Backus-Naur Format (BNF), which had been used to describe command languages, provided a uniform and well-defined metasyntax that could be used to compare various languages. One problem with BNF was that it could not capture some language uniformities that would be

obvious to users. Thus BNF could not describe the psychological complexity of a command language. This was unsatisfactory because Reisner hoped to not only to describe the structure of the language, but also predict the psychological complexity of the language. Reisner (1981) described a formalism that predicted empirical complexity. Others such as Moran (1981) and Kieras and Polson (1985) also attempted to specify the complexity of user interfaces. Card, Moran and Newell (1983) developed an elaborate Keystroke-Level Model that predicts performance times of expert users.

This approach was advanced when Payne and Green (1986) proposed a formalism that they called a task-action grammar. This formalism attempts to map tasks onto actions in such a way that psychologically real aspects of the interface are captured and predictions can be made about learnability. Payne and Green (1986) reported experiments that showed that TAG correctly predicted the comparative learnability of two languages carefully constructed such that a surface (keystroke level) analysis yielded the opposite prediction.

Measuring Learnability

Learnability is a function of inconsistency in a command language. Payne and Green were concerned with measuring the consistency of action languages on three levels: lexical, syntactic, and semantic. Lexical inconsistency results from the lack of congruence between the command language lexemes and their external meaning. For example, if the command to go to the beginning of a file is UP then it would be inconsistent to use FORWARD to go to the end of a file because DOWN would be expected.

Syntactic inconsistency results from conflicts in the organizing principles of the language. If, for example, the option key is used to enhance the effect of another key in some situations, but it is used to constrain actions in other situations no single principle can be attached to the option key. Additionally, if the order of operators and their arguments vary, this can be a kind of syntactic inconsistency. In this regard, Barnard, Hammond, Morton and Long (1981) found that positionally consistent systems were most readily learned.

Semantic inconsistency results from incompleteness in the functionality of the language. If one can search both forward and backward, one would expect to be able to search and replace in both directions. The lack of such a function would be a flaw in the language and would lead to user errors.

Payne and Green's TAG is a context-free generative grammar that maps tasks (semantic units) onto user actions (lexical units). A generative grammar is simply a set of rules that defines the structure (syntactic units) of a string of symbols. Rules consist of a left-hand side (LHS) and a right-hand side (RHS). In a context-free grammar only one term can appear on the LHS. The RHS can consist of terms or terminal symbols. A terminal symbol is the equivalent of a user action.

A grammar consists of a dictionary and a set of rule-schemas. The dictionary is a list of the simple tasks that the user can perform routinely. A task consists of a label and a set of features that categorize it uniquely. The following are examples of dictionary entries:

jump-to-label{Effect = jump, Place = +label}

jump-to-beginning{Effect = jump, Place = -label, Direction= backward}

jump-to-end{Effect=jump, Place = -label, Direction =forward}

The labels given to the tasks are arbitrary and redundant. In all cases the label could be replaced with Task. However, the label has a mnemonic value and should be chosen to be meaningful. In the above example the three tasks are distinguished by two features, Place and Direction, each of which may take two values. A single dimension with three values would have served as well, although that change would require the grammar to take a slightly different form.

Below is an example of a rule-schema. The semantic features of one of the above tasks would be inserted into the rule-schema:

Task[Effect=jump, Place]. "J" + select-place[Place]

The LHS consists of generic label and two features, one of which is already valued. The fact that the feature Effect is already set to jump blocks the application of any other tasks with other effects. The RHS consists of the terminal symbol "J" that indicates the user must type J. The other term select-place[Place], since it is non-terminal, becomes the LHS of another rule and it is evaluated with Place set to the correct value. This process continues until only terminal symbols remain.

The initial motivation for writing such a grammar was to assess the learnability of different comparable languages. A TAG description makes the empirical prediction that the language with the fewer simple-task rule-schemas will be easier to learn. If two languages are equivalent in this respect then the language with fewer rules overall will be easier to learn. This is useful information for instructional designers because we have few formal ways of measuring learning difficulty, but the primary use of such a grammar is to capture at an abstract level the rule-schemas that a skilled user will have mastered. In other words, instead of conceiving of rules as isolated units, this technique allows us to represent rules at an abstract level that has been shown empirically to make good predictions about learnability. This abstract representation opens the door to an alternative instructional strategy. Conventional prescriptions suggest that rules should be taught at a very concrete level, proceeding from simple and frequent to complex and infrequent. A TAG suggests that teaching at an abstract level (the rule-schema rather the rule) might be more efficient and effective since it is precisely this abstract representation that the student must learn to master the command language.

D-TAG

The original TAG notation was modified according to Howes and Payne (1990) to include two aspects of devices. The first aspect is the information carried by the display. Because a display provides some of the information needed to perform a task, this feature should be incorporated into a TAG. The second aspect is the side-effects of an action that are not the direct object of the action, but nonetheless are manifested as a result of the action.

The display feature in the TAG is motivated by the fact that users often cannot remember the actual configuration of displays but can recognize needed functions when they are visible. This recognition process is facilitated if the task semantics match the display semantics. This fact is captured with the notation display_item(<task features> <display features>). Task features are derived from the task dictionary. Display features are a set of features (i.e., variables) which allow the display item to be located.

Side effects are represented as rewrite rules that specify an object in the LHS and a change of value for some feature of the object in the RHS. Thus a side-effect rule might read side_effect(Obj=button) --> effect(value_from_display_item <-- highlight). When that rule is executed it will be evaluated to some visible or invisible change of status of some aspect of the machine and expert users will be able to update their internal model of the device.

To make actions more explicit the notation was modified (Howes & Payne, 1990) to represent all actions as: action("action","object"). Thus the action of clicking on a button located on the upper-left-hand side of the screen would be represented as action(press, Display-item([hot-spot], class=button, location=upper-left).

With these changes a D-TAG was written to represent the expert user's model of the Casio command language.

Future Plans and Conclusions

The next logical step with the simulator and the TAG is to perform experiments testing various approaches to instructing novices in the operation of the watch. The conventional approach, as described above, is to teach rules at the literal action level. Even this low level prescription allows some leeway for interpretation, because the ordering of the rules and even the ordering of the steps is subject to differing hypotheses about the optimal approach. Thus an experiment comparing various literal action level approaches with an abstract rule-schema approach would generate some interesting data. Current instructional design theory does not mention abstraction as an approach to rule learning. Thus the simulator and the TAG provide a rich practical and theoretical basis for enhancing instructional design theory.

References

Barnard, P. J., Hammond, N. V., Morton, J., & Long, J. B. (1981). Consistency and compatibility in human-computer dialogue. International Journal of Man-Machine Studies, 15, 87-134.

Brown, J. S. (1986). From cognitive to social ergonomics and beyond. In D.A. Norman & S.W. Draper (Eds.), User-centered system design (pp. 457-486). Hillsdale, NJ: Lawrence Erlbaum Associates.

Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum.

Gagné, R. M., Briggs, L. J., & Wager, W.W. (1992). Principles of instructional design (4th ed.). New York: Harcourt Brace Jovanovich College Publishers.

Gustafson, K.L., & Tillman, M.H. (1991). Designing general strategies of instruction. In L.J. Briggs, K.L. Gustafson, & M.H. Tillman (Eds.), Instructional design principles and applications (pp. 173-192). Englewood Cliff, NJ: Educational Technology Publications.

Howes, A., & Payne, S. J. (1990). Upper LCD-based competence: Towards user models for menu-driven interfaces. International Journal of Man-Machine Studies, 33, 637-655.

Jonassen, D.H., Hannum, W.H., & Tessmer, M. (1989). Handbook of task analysis procedures. New York: Praeger.

Kieras, D., & Polson, P. G. (1985). An approach to the formal analysis of user complexity. International Journal of Man-Machine Studies, 22, 365-394.

Leshin, C. B., Pollock, J., & Reigeluth, C. M. (1992). Instructional design strategies and tactics. Educational Technology Publications: Englewood Cliffs, NJ.

Moran, T. P. (1981). The command language grammar: a representation for the user interface of interactive computer systems. International Journal of Man-Machine Studies, 15, 3-50.

Okey, J.R. (1991). Procedures of lesson design. In L.J. Briggs, K.L. Gustafson, & M.H. Tillman (Eds.), Instructional design principles and applications (pp. 193-208). Englewood Cliff, NJ: Educational Technology Publications.

Payne, S. J., & Green, T. R. G. (1986). Task-action grammars: A model of the mental representation of task languages. Human-Computer Interaction, 2, 93-133.

Payne, S. J., & Green, T.R,G. (1989). The structure of command languages: An experiment on task action grammars. International Journal of Man-Machine Studies, 30, 213-234.

Polson, P. G. (1987). A quantitative theory of human-computer interaction. In J. M. Carroll (Ed.), Interfacing thought: Cognitive aspect of human-computer interaction. Cambridge, MA: MIT Press.

Reisner, P. (1977). Use of psychological experimentation as an aide to development of a query language. IEEE Transactions on Software Engineering, SE-3, 218-229.

Reisner, P. (1981). Formal grammar and human factors design of interactive graphics system. IEEE Transactions on Software Engineering, SE-7, 229-240.

Smith, P. L., & Ragan, T. J. (1993). Instructional design. Merrill: New York.

Appendix 1.

<u>D-TAG for Casio Watch</u>

Task Features

mode   dateTime, altitude, barometer, depth, alarm, countDown, stopWatch

operation   yes, no

runMode   stop, run, split, stopsplit

operation   yes, no

direction   increment, decrement

Task   Dictionary

selectMode[mode=dateTime]

selectDateTimeMode[mode=altitude,barometer,depth,alarm,countDown,stopWatch; operation=no]

selectDateTimeMode[mode=altitude, barometer, depth, alarm, countDown, stopWatch; operation=yes]

Set[mode=alarm, altitude, countDown, depth, dateTime]

Set[mode=barometer]

altDisplay[mode=altitude,depth]

altDisplay[mode=dateTime]

timer[mode=stopwatch,countDown; runMode=stop]

timer[mode=stopwatch,countDown; runMode=run]

splitRelease[mode=stopwatch; runmode=split]

displayStopSplit[mode=stopwatch; runmode=stopsplit]

setAlarmBeep[mode=alarm]

Rule-schemas

Convention: **action(press, Display-item([button], location=lower-left))** *
means repeat until target display is reached. This can be expressed by a set of rules
schemas, as illustrated below.

•Task[effect=selectMode, mode=dateTime]

-> action(press, Display-item([button], location=lower-left)) *

•Task[effect=selectDateTimeMode, mode=[not dateTime]; operation=yes]

-> action(press, Display-item([button], location=lower-left)) +sideEffect(operation)

•Task[effect=selectDatetimeMode, mode=[not dateTime]; operation=no]

-> action(press, Display-item([button], location=lower-left)) *
+sideEffect(displayValue)

•Task[effect=Set, direction= :decrement,
mode=dateTime/altitude/depth/countdown/alarm, operation]

-> begin + action(press, Display-item([button], location=lower-left)) * +
action(press, Display-item([button], location=upper-right)) * + end
+sideEffect(operation)

•Task[effect=Set, direction= :increment,
mode=dateTime/altitude/depth/countdown/alarm, operation]

-> begin + action(press, Display-item([button], location=lower-left)) * +
action(press, Display-item([button], location=lower-right)) * +
end+sideEffect(operation)

•Task[effect=Set, direction= :decrement, mode=barometer, operation]

-> beginBar + action(press, Display-item([button], location=lower-left)) * +
action(press, Display-item([button], location=upper-right)) * +
end+sideEffect(operation)

•Task[effect=Set, direction= :increment, mode=barometer, operation]

-> beginBar + action(press, Display-item([button], location=lower-left))* + action(press, Display-item([button], location=lower-right))* + end + sideEffect(operation)

•Task[effect=displayMax, mode=altitude/depth]

-> action(press, Display-item([button], location=lower-right))

•Task[effect=toggleDateTime, mode=dateTime]

-> action(press, Display-item([button], location=lower-right))

•Task[effect=start-timer, mode=stopwatch/countDown, runmode=stop, operation]

-> action(press, Display-item([button], location=lower-right)) + sideEffect(operation)

•Task[effect=reset-timer, mode=stopwatch/countDown, runmode=stop]

-> action(press, Display-item([button], location=upper-right))

•Task[effect=stopTimer, mode=stopwatch/countDown,runMode=go]

-> action(press, Display-item([button], location=lower-right))

•Task[effect=splitTime, mode=stopwatch,runMode=go]

-> action(press, Display-item([button], location=upper-right))

•Task[effect=splitRelease, mode=stopwatch,runMode=split]

-> action(press, Display-item([button], location=upper-right))

•Task[effect=splitStop, mode=stopwatch,runMode=split]

-> action(press, Display-item([button], location=lower-right))

•Task[effect=displaysplitStop, mode=stopwatch,runMode=stopsplit]

-> action(press, Display-item([button], location=upper-right))

•Task[effect=setAlarmbeep, mode=alarm, operation]

-> action(press, Display-item([button], location=lower-right))*
+sideEffect(operation)

begin[mode=dateTime/altitude/depth/countDown/stopwatch/alarm]

-> action(press, Display-item([button], location=upper-left))

end[mode=dateTime/altitude/barometer/depth/countDown/stopwatch/alarm]

-> action(press, Display-item([button], location=upper-left))

beginBar[mode=barometer] -> action(press, Display-item([button], location=upper-left)&Display-item([button], location=lower-left))

[for simulator this is rendered as press b & shiftkey simultaneously]

sideEffect(operation=no)

-> effect(operation <- yes)

sideEffect(operation=yes)

-> effect(operation <- no)

The following shows how an action can be called repeatedly until the desired display is achieved.

action(press, Display-item([button], location)*

LCD-item([text | symbol | number],form≠value-from-goal)

->

action(press, Display-item([button], location))

side effect(LCD)

action()*

action()*

LCD-item([text | symbol | number], form=value-from-goal)

->

action(null)

side-effect(Upper LCD, value-"null")-->

effect(Upper LCD, label="ALT")

side-effect(Upper LCD,value-"ALT")-->

effect([Upper LCD], label="ATM")

side-effect(Upper LCD,value-"ATM")-->

effect([Upper LCD], label="DEPTH")

side-effect(Upper LCD,value-"DEPTH")-->

effect([Upper LCD], label="ALM")

side-effect(Upper LCD,value-"ALM")-->

effect(flash Lower LCD, label="CD")

side-effect(Lower LCD,value-"xx:xx xx")-->

effect([flash Lower LCD], label="SW")

side-effect(Lower LCD,value-"xx:xx xx")-->

effect([Upper LCD], label="dateTime")

AUTOMATED DETECTION OF INDIVIDUAL RESPONSE CHARACTERISTICS
IN TRACKING TASKS: AN EXPLORATORY STUDY USING NEURAL NETWORKS

Ryan D. Tweney
Professor
Department of Psychology
&
Susan T. Chitwood
Research Assistant
Department of Psychology

Bowling Green State University
Bowling Green, OH 43403

AUTOMATED DETECTION OF INDIVIDUAL RESPONSE CHARACTERISTICS

IN TRACKING TASKS: AN EXPLORATORY STUDY USING NEURAL NETWORKS

Ryan D. Tweney & Susan T. Chitwood
Department of Psychology
Bowling Green State University
Bowling Green, OH 43403

Abstract

The possibility of detecting individual differences in a complex tracking task was examined. Tracking data from a dual task study designed to examine the effects of workload on performance was examined. Neural networks were used to characterize individual subject performances. A variety of network types and architectures were explored, in an attempt to determine: (a), whether differences in performance configuration among different subjects could be detected by the networks, and, (b), whether the networks could discriminate the specific experimental condition that a particular individual record of performance scores had been gathered in. Success in both tasks was shown to be highly dependent upon the nature of the data used for analysis. In general, networks can detect some individual characteristics from raw performance data, and, in some circumstances, may be able to characterize the external conditions that produced particular response configurations. Extensions of the present methodology to performance data gathered in real-time flight simulations thus seem potentially promising and are under consideration for future investigation.

AUTOMATED DETECTION OF INDIVIDUAL RESPONSE CHARACTERISTICS IN TRACKING TASKS: AN EXPLORATORY STUDY USING NEURAL NETWORKS[1]

Ryan D. Tweney & Susan T. Chitwood

Attempts to understand the nature of human performance characteristics in complex task situations characterized by varying levels of workload have, in recent years, been seen as critical to a full understanding of the nature of situation awareness (SA). According to Endsley, SA can be defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (1991, p. 7). From this definition, the relation to workload issues can be appreciated: If SA depends upon correct perception and comprehension, as well as upon inferences to future states of the system, then it follows that factors which interfere with such performance can be deleterious to SA. In particular, Endsley (1991, p. 41) identified attentional effects as important consequences of high workload: when workload increases, the narrowing of attention that frequently follows is of potentially great concern.

One methodology used to examine such issues involves the performance of subjects in laboratory tracking tasks under varying conditions of time and/or accuracy demand, sometimes including multiple task workload manipulations as well. Wickens (1984) noted that such studies had not adequately explored performance on multiple tasks, generally assessing performance only on the tracking task (considered to be of primary interest) and ignoring the need to characterize the effects on all tasks that the subject was engaged in. To address this lack, Vidulich (1989; 1990) conducted a study in which a priority manipulation was included. Subjects were required to carry out both continuous tracking and a short-term memory monitoring task (the "Sternberg Task") and were instructed to consider one of the two tasks as primary, and to focus upon that one, while continuing to perform the other. The results suggested that the procedure of manipulating priority was, in fact, a sensitive method for

---

assessing the effects of workload on performance, and possessed implications for theoretical models of the nature of performance in multiple-task situations.

One barrier to the full understanding of workload effects on SA centers on the likely presence of large individual differences in response to increasing workload. While it is clearly true, in general, that performance generally declines as workload increases, it is also true that the rate, topography, and asymptotic characteristics of that decline vary enormously from one individual and/or situation to another. This is not, of course, surprising: individual differences are large in any complex task, and the situations of most interest in the SA literature are complex.

In the present series of studies, the nature of individual differences in data derived from Vidulich's (1990) study was explored using neural networks. Since much of the individuating characteristics of tracking data can reasonably be expected to be nonlinear in character, it was felt that neural networks, which are able to detect such structure, could prove to be an effective way to isolate and examine their character. Specifically, we asked two questions: (1) Can a neural network, if given tracking data from an individual subject running under different experimental conditions, learn to correctly identify the condition ? (2), Can tracking performance data from two or more different subjects be discriminated by a neural network? The first question presupposes that the particular individuating characteristics of a subject's response to the experimental manipulation are captured by the tracking performance data, while the second question presupposes that individual differences are present in the data. Neither question presupposes anything about the shape of those differences. Further, the first question can be seen as a kind of "strong test" of any generalizations reached on the basis of an overall statistical analysis of the group averaged data -- if a neural network can discriminate the experimental condition of a particular subject based only on that subject's response patterns, then the experimental manipulation has had an effect on the subject in question. In theory, then, one ought to be able to separate the subjects in an experiment into three groups:

> (a) subjects that confirm an overall statistical effect, because they are learnable by neural nets and
>
> > manifest the same pattern present in the group data,
>
> (b) subjects that disconfirm the overall statistical effect, because they are learnable by the neural nets but
>
> > manifest a different pattern than that present in the group data,

(c) subjects that provide no evidence toward the group effect, because they cannot be learned by neural nets.

Statistically significant effects due to an experimental manipulation do not, of course, dictate the number of subjects that will fall into one or another of the above categories. Nonetheless, inspection of the numbers within each classification are potentially of great value as a supplement to traditional analysis -- an obtained effect will need to be interpreted very differently if all of the subjects fall into the first of the three classes than if some fall into the second and/or third class! In effect, the neural network analysis can establish the extent to which a given statistically significant empirical generalization is true of all of the subjects in an experiment or of only some of the subjects.

In recent years, neural networks have been applied to a large variety of practical and theoretical questions -- in pattern perception, speech recognition, signal processing, and so on (see the papers in Mehra & Wah, 1993, for examples.) In effect networks can be regarded as "self-programming" systems that configure themselves, according to some criterion, on the basis of their experience with the "external world" manifested in data inputs to the network. In the present application, we are not providing an a priori model to the network; if it is to discriminate, say, the experimental condition of a subject, it will be necessary for the network to detect some individuating characteristic of the data set. Thus, the network's ability to capture highly configural characteristics is important here; they can potentially serve as discovery procedures for aspects of data not revealed by traditional, model-based, statistical analyses.

Networks are composed of a number of interconnected processing elements, or "nodes," which can take on differing activation levels (Rogers, Kabrisky, Ruck, & Tarr, 1990.) Via the interconnections, the activation of any one node is a function of the activation of other nodes with inputs to that node and, in turn, has an influence on the activation of other nodes to which it is connected. The connecting links between nodes are provided with "weights," numerical values that scale how much (or how little) of the activation of a source node is conducted to a target node. For any one node, the total input is generally summed by taking the (weighted) outputs of all the source nodes, and passing this value through an activation function (generally a sigmoidally-shaped function, such as a logistic curve or a hyperbolic tangent) to determine the activation level of the target node. Networks can be "trained" in a supervised procedure by starting with an initial random guess as to the output for each input and

receiving feedback, usually in the form of a discrepancy between the desired output and the actual output. The discrepancy is then usable as an index of how much (and in which direction) the existing interconnection weights need to be changed. As this procedure is repeated in multiple training passes through the data sets, the weight changes can converge on a stable set of values that, ideally, give the desired output associated with each input.

The most frequently employed neural network configuration employs three or more layers of processing elements trained using one or another variant of the backpropagation learning algorithm (Rumelhart & McClelland, 1986). Finding a specific training procedure that produces the right network weights for any given problem, however, is still largely a matter of skill: no general algorithm exists that is guaranteed to find an optimal combination of weights within an arbitrarily specified time. As a result, a number of heuristics have evolved for training neural networks. These heuristics are of sufficient power to enable surprisingly complex problems to be solved. In the present study, a variety of such heuristics were employed in attempting to find the optimal neural networks for a given data set.

Some pilot data was available to suggest that using networks to classify tracking data could be carried out (G. Reid, personal communication, July, 1992). Accordingly, a systematic investigation of this possibility was undertaken, using tracking data from the Vidulich experiment, and employing a variety of neural network approaches to insure that optimal network solutions were achieved.

Methodology

The tracking study used as a source of data has been described in detail elsewhere (Vidulich, 1990; 1994). The relevant aspects of the task will be briefly described here, followed by a general description of the neural network approach used. The specific details of the separate phases of the neural network studies will be described at the appropriate places in the Results section of this report.

The Tracking Task: As noted earlier, the tracking data used in the present study were derived from a task situation in which subjects were required to perform two task simultaneously, a continuous tracking task and a Sternberg memory task. Data from 21 subjects (11 males and 10 females), ranging in age from 19 to 29, were gathered over a period of five days in sessions lasting from 1.5 to 2.5 hours. Stimuli for both tasks were presented on a monitor in front of the subject, the Sternberg task stimuli being presented in a box located above the tracking task on the screen. The tracking target was a 1/4" x 1/4" cross centered on the screen. A forcing function kept a

vertical line cursor, 1/4" high, in motion along a horizontal line passing through the target. The subject's task was to keep the cursor over the centering cross. Cursor movement was controlled by a joystick located to the subject's right. During an experimental trial, the subject attempted to offset the effect of the forcing function by continuous compensatory movements of the joy stick. At randomly chosen times within the trial, the Sternberg task was presented. Subjects responded to this task by means of button presses using their left hand. Instructions emphasized that either the Sternberg or the tracking task was the highest priority. Tracking data were sampled every 33 msec. The raw data were divided into segments, corresponding to 12 "windows" 33 msec in width, surrounding each Sternberg event. For the tracking task, during each window interval "Control Movement" of the joystick was recorded (actually, a measure of the average velocity of the stick during that interval), as well as the position of the cursor. For each window, further measures of tracking performance were derived from the Control Movement data by incorporating the concurrent values of the forcing function and cursor location (a "Control Effectiveness" score) and the cursor location and target location (a "Root Mean Square Error" score, RMSE). Note that RMSE data and Control Effectiveness data change continuously regardless of subject input, since the forcing function keeps the cursor in motion and can bring it near to the target in spite of the subject's behavior (or lack of behavior!) As will be seen, this is an important aspect of the measures. In particular, neither permits the direct examination of hesitations, times when the joystick control movements cease. Such hesitancies may be relevant in accounting for performance on tracking tasks (Netick & Klapp, 1994).

The experimental design included manipulation of Sternberg memory set size (2 or 4 elements), complexity of response in the Sternberg task (a single key or a three-key response for "Yes" or "No" responses), and tracking order (a first order, velocity control, task vs. a more difficult second order, acceleration control, task). The manipulation of most concern to the present study was task priority. During the 4th and 5th experimental session, subjects were instructed to give priority to either the Sternberg task or the tracking task. Priority order was counterbalanced across the two sessions. In the present study, most of our attention was restricted to the first order tracking task, Sternberg set size of 4, Sternberg response complexity of 3 (i.e., the most difficult condition for the Sternberg task, and the easier of the two tracking tasks). Analysis of variance of the results for this combination of conditions alone (i.e., for the 2 x 12 design of Priority Level by Window) was provided to us by Mike Vidulich (personal communication, 1994) for two of the dependent variables (RMSE and Control Movement). There was an

interaction between Priority Level and Window for both variables. There were main effects due to Priority Level and Window for RMSE but due to Window only for Control Movement data.

Neural Network Methodologies: All neural networks were implemented in "NeuralWorks Professional II+ (Version 5.0)," a commercial neural network software package (NeuralWare, Inc., 1993), running on either 386 or Pentium platforms. Data files were constructed from the original data files (described in Vidulich, 1994), edited in Microsoft Excel into the appropriate format for NeuralWorks II+, and converted to ASCII files. Each file contained all of the observations for one subject in two of the experimental conditions, Tracking Priority vs. Sternberg Priority (with other conditions held constant as noted above). Each file contained approximately 60 lines of data for the Tracking Priority, and approximately 60 lines for the Sternberg Priority. Each line contained the 12 observations made within each window surrounding a Sternberg event (which was always located in the fourth window). Priority level was coded at the end of each line. Each subject file contained approximately 120 cases available for training a neural net, roughly 60 from each condition.

In all work with neural networks, an important question concerns just what a network has learned. Existing algorithms are so powerful that they can, in effect, learn "anything," even a random configuration of patterns, if provided with a sufficient number of hidden units. This problem, sometimes known as "Overfitting" (Masters, 1993), occurs because the number of free parameters in a network (in the form of modifiable connection weights) can be used to simply memorize each training case presented to the network. To avoid the problem, it is necessary to use a crossvalidation procedure, in which the network is trained initially on a randomly chosen subset of the training cases, and then tested on the remaining cases. If the network has merely memorized training cases, then it will perform poorly on the test set. On the other hand, if the network has extracted general features from the training set, then it is likely that these features will allow it to perform well on the test set. Defeating the problem of overfitting is a major part of the "art" of neural networks. If a given network learns the training classification but fails to generalize, then one strategy is to reduce the number of free parameters in the network by removing hidden units. However, this can also have the effect of making learning more difficult or impossible. Thus, the trick is to develop generalizable networks that can be trained within moderate periods of time.

In most cases, the networks we trained had an input layer consisting of 12 nodes (one for each window in a data line) and one or two output nodes. Initially, we attempted to train two output nodes, corresponding to the

1-0 or 0-1 coding used in the data files. However, strictly speaking, the problem requires only one output node, which can capture the binary classification by varying from 0 to 1, a coding which was used on the later networks in the study.

NeuralWorks II+ provides a number of diagnostic aids for assessing the performance of a particular network. As a network is being trained, the network's output for each input vector is compared to the intended output, and Root Mean Square Error is computed and averaged across all training cases (RMS, here, is not to be confused with RMSE, the dependent measure of performance in the tracking task). NeuralWorks II+ then plots a running record of RMS error which can be used to monitor the performance of the network. In addition, NeuralWorks II+ provides a measure of "Classification Rate," an index of the percentage of hits, misses, false alarms, and false positives associated with the network's outputs. These values served as the primary index of network performance, rather than the RMS measure. For our purposes, all that counts is that a particular output be greater than 0.5 (hence predictive of a "1" in the training file) or less than 0.5 (predictive of a "0"); such performance is better captured by the Classification Rate than by the RMS value. In general, we report only the proportion of hits given by the Classification Rate analysis, comparing these across the training and test sets.

Results

In general, we found that the experimental condition (Sternberg Priority vs. Tracking Priority) could be learned from the data of a single subject, but only when Control Movement (CM) data was used, and only for some of the subjects. We were not able to find a generalizable solution when RMSE data were used, in spite of a large number of attempts. In this section, we first describe the analyses of the RMSE data, and then the CM data.

I. Analyses Using RMSE Data: In our initial attempts, RMSE data was obtained from the second order tracking task. Since this task requires that the subject adjust the acceleration of the cursor by moving the joystick, it is very much more difficult than the first order velocity control task. While ANOVA of the experimental results showed a significant main effect due to task priority, we were not able to find a neural network that could discriminate the experimental condition from the data of a single subject. However, the difficulty level of the task is so high that this was not seen as fatal to the overall goal of the project. In effect, when RMSE values are so large, they may not reflect the workload variables of most interest. The decision was therefore made to abandon further analysis of this data, and to focus our remaining efforts on first order tracking data.

<u>Study 1</u>: Data from six subjects was chosen for analysis. One of the six (#44) had been used in the pilot study and had appeared then to be learnable and generalizable. The other five were chosen arbitrarily. Each was tested with six neural network architectures and a variety of learning schedules. All of the networks used two output units, scaled from 0 to 1, to represent the classification into Training vs. Sternberg Priority, and all of the networks used an activation rule based upon a sigmoidal function of the form activation = $1/(1 - e^{-x})$. Two of the networks had a single hidden layer with three processing elements, but differed in that two different learning algorithms were employed, a standard backpropagation routine and a "Delta-Bar-Delta" (DBD) routine developed by Jacobs (1988). In the latter, the learning rate of specific elements is adjusted by a routine that looks at the recent history of weight changes for that element -- if the weight changes are of alternating sign (i.e., if the local gradient region is heavily curved, and the weight change has been oscillating), then the overall learning rate for that element is reduced. DBD networks have been shown to be better at avoiding slow convergence of a network under the usual backpropagation algorithms. In addition to the single hidden layer networks, four networks with two hidden layers were used, one with 21 and 11 units (replicating a successful network from the pilot study) and three networks with 10 and 2 hidden units, each using a slightly different learning schedule. "Learning schedule" in this context refers to a systematic change in the learning rate over trials, reflecting the fact that high learning rates are best used at the beginning of training but can produce excessive oscillation once the network is partially trained. It is, therefore, standard procedure to allow the learning rates to decline in a stepwise fashion as training proceeds.

Of the 36 networks run in Study 1, some were able to learn the training data quite well (Proportion of correct classifications = .90 or better), but none was able to generalize; for most, performance on the test set was no better than .60 correct.

Since all of the above networks were based upon one random sort of the data from each subject (i.e., one random cut into a training and a test set), we decided to sample a variety of such cuts, to insure that our results were not biased by sampling variation. Accordingly, data from one subject (#44) was randomly divided into training and test sets in 7 different random fashions. Various configurations of single and double layers and of learning schedules were used. In all, approximately 140 networks were tested. While some could learn the training set nearly perfectly, none was able to generalize to the test set with greater than 60% correctness.

Study 2: The failure of Study 1 led us to ask whether there was a way to "smooth" the data by taking advantage of a Self-Organizing Network, or "Kohonen Net," to pre-process the data (Kohonen, 1990). Such a net is a rectangular array of processing elements which "compete" with each other to be the only unit representing a specific input training case. No correct example is provided to a Kohonen net. Instead, training cases are fed repeatedly to the network, and processing elements modify their own weights as they compete to represent specific units. At the end of learning, each training input activates one and only one processing element, but any one processing element can represent more than one training input (if those inputs are similar to each other). Kohonen nets thus remove some of the "noise" in a training set. After Kohonen training, the outputs from the Kohonen units can be backpropagated in the usual fashion (i.e., in a supervised learning procedure) to generate the needed classification into Training vs. Sternberg Priority.

We built 19 Kohonen nets using subject #44. The nets varied in size of the Kohonen layers, ranging from 25 units in a 5 x 5 array to 64 units in an 8 x 8 array. Learning schedules were varied in the backpropagation phase of training. In contrast to Study 1, which used two output units varying from 0 to 1, in this study, two output units varying from 0 to 1 or from 0.2 to 0.8 were used. The latter value is probably more appropriate since the output of a sigmoid function is asymptotic at 0 and 1. Thus, a network can never learn a 0 to 1 classification perfectly.

A number of the 19 nets were able to learn the final classification in the .8 to .9 range. None was able to generalize appropriately, however -- the highest generalization obtained was 0.68.

Study 3: In all of the studies thus far, the training and test sets were developed by taking 50% of the total data available for each set (roughly 60 cases in each). In Study 3 we explored the possibility that too little training data was being made available for generalizable learning to occur. Accordingly, a larger sample was taken from seven of the subjects, using 75% of the data for training, and 25% for testing. Five network architectures were chosen (based on the most promising networks run in Study 1), for a total of 35 networks. Initial training led to classification rates ranging from 0.50 to 0.99, but generalization was not apparent: classification rates ranged from 0.42 to 0.81, the latter being observed in only one case.

In a variant of the preceding study, six subjects were used in a 25% - 25% cut of the available data, with the thought in mind that the unused cases could serve as a further test, if generalizable learning was found with the

smaller training sets. Such sets are apparently too small, however. Of the 30 networks tested (6 subjects x 5 network configurations) many learned during training in the 0.9 range or better, but generalization was usually in the 0.5 to 0.6 range.

Study 4: The preceding studies used two output units. However, one output unit should be sufficient to represent a binary classification. Further, since one output unit reduces the number of connections by half, this should make memorization harder. Study 4 was conducted by using one hidden layer (3, 4, or 5 hidden units), a single output node (scaled from 0.1 to 0.9, 0.2 to 0.8, or 0.3 to 0.7), and two different leaning schedules, for a total of 18 networks. Data from subject #44 were used.

Some of the networks learned the classification to better than 90% accuracy, but none generalized beyond 0.80. It was observed that the better the learning, the poorer the generalization, with optimal performance seeming to be at about 0.80 learning and 0.70 or so generalization.

Study 5: In a final effort to determine if the classification into Tracking Priority vs., Sternberg Priority could be learned and generalized, data from subject #44 was subjected to a variety of networks. These attempts will not be detailed here, but included Cascade Learning, use of a sine activation function, use of a hyperbolic tangent activation function, varying numbers of hidden units in two hidden layers, and an "undertraining" procedure, in which networks are stopped early in training, presumably blocking the possibility of memorization. Cascade Learning (Fahlman, 1990) seemed especially promising. This is a procedure in which a network is built with no hidden units and trained until no further improvement in error is noted, after which one hidden unit is added and the network trained again until no further error improvement is noted, and so on. In effect, this procedure should insure that maximum generalizability is obtained because it should be able to find a solution with the minimum number of hidden units necessary. None of the procedures succeeded, however, not even the Cascade Learning procedure. While some training results in the .90s were obtained, generalization was poor to mediocre, with the best overall solutions seeming to be a training classification rate of about 0.80 and a test classification rate of about 0.80 as well.

Study 6: Our final study using RMSE data employed a procedure suggested to us by Major E. Fix (personal communication, August, 1994). In this procedure, a "data encoder" is first built, that is, a network which takes the training cases as input and is required to produce the same vector as output. By limiting the number of

hidden units to a smaller number than the number of components in the training vectors, one guarantees that the network cannot simply memorize each case. Thus, for our data, a 12-3-12 encoder could be built, with the number of hidden units being too small to permit memorization to occur. Once trained to reproduce the training set, the output layer is cut off, the weights connecting the input layer to the hidden units are frozen, and training carried out again by connecting the hidden units to the single output node representing the target classification.

Initial testing showed that using only 1, 2, or 3 hidden units made the encoding problem insoluble. With 4 or more hidden units, the encoding was successful. The results of second-stage training for 4 through 11 hidden units (using, as before, the single subject from the pilot study) were as follows:

Number of Units	Training Classification	Test Classification
4	.78	.76
5	.86	.56
6	.89	.60
7	.92	.83
8	.95	.63
9	1.00	.48
10	.95	.70
11	1.00	.60

Note that these results show increases in initial learning as the number of hidden units increase, but that generalizability peaks at about 7 hidden units. After that, some degree of memorization appears to take place and generalization scores appear to drop.

Summary of Studies 1 to 6: All of the above studies used RMSE data and all failed to produce a generalizable solution for any of the subject data sets. Inspection of the data sets suggests that a great deal of variation is present, and it may be that asking a neural net to learn general properties in very noisy data is an inappropriate application. Analysis of variance, by contrast, is especially adapted to such a problem -- noise is removed by averaging across subjects and estimating the noise component of the data from the separate variabilities of cells within the design. This permits a sensitive look at otherwise hard-to-detect patterns, though it can only be used when a prior model is specified and data averaged across cells. The intention in our study was

different. We hoped to detect subject-specific, possibly highly configural, patterns, without specifying in advance an underlying model that specifies what those patterns could be. Perhaps this cannot be done in the presence of excessive noise, which both masks configural patterns and provides a rich (and spurious) set of arbitrary patterns which the network can memorize.

A second consideration concerns the source of noise in the RMSE data. Such data is not, in the final analysis, purely a behavioral measure. Instead, as noted earlier, it is a derived score that depends upon what the subject does <u>and</u> upon what the cursor does. Since the latter is driven by a forcing function, the RMSE is a product of two potentially very different processes. In the usual analysis of tracking data, this does not matter because RMSE is measured across fairly wide intervals, over which the forcing function can be expected to cancel itself out. The same may not be true for small-window averages such as those in the present data set (M. Vidulich, personal communication, August, 1994). Accordingly, we decided to examine data that was more closely tied to the subject processes that are of principal concern.

<u>Study 7 -- Analysis Using CM Data</u>: Control Movement (CM) data, a real-time measure of the velocity of the control stick, was chosen from 5 subjects. One network was built for each, using 3 hidden units, one output unit (scaled from 0.2 to 0.8), a sine activation function, and regular backpropagation. All 5 networks learned the training inputs readily, correctly classifying the Task Priority manipulation in well over 90% of the cases. Further, the generalization results were extremely promising as well: the test set results were 0.91, 0.84, 0.82, 0.60, and 0.57, respectively. The first value, 0.91, is the highest observed in any of the prior studies, and the second and third are within striking distance of acceptability. The result is all the more remarkable because no special attempt was made to optimize network performance.

<u>Study 8 -- Subject Discrimination Using CM Data</u>: Since the networks seemed to be able to discriminate the experimental condition of the subjects, five subjects were chosen at random, and an attempt was made to use the networks to discriminate one subject from another on a pairwise basis. Inputs consisted of the 12-position input vector, as before, but training sets were devised by combining Control Movement data from two different subjects. The networks were trained to discriminate subject identity. All networks had 3 hidden units and one output unit (scaled from 0.2 to 0.8), used a sine activation function, and a default learning schedule. We were able to complete the pairwise comparison of Subject 21 with four others, and the results were as follows:

Comparison	Training	Test
21 vs. 22	.92	.68
21 vs. 36	1.00	.27
21 vs. 42	.86	.53
21 vs. 44	1.00	.94

Note that Subject 21 and 44 are, by this criterion, learnably distinct, but that 21 and 36 are not. In effect, such results could be used to combine subjects into groups whose overall response topography is similar within groups (e.g., 21 and 36) and others who must be regarded as different in character (e.g., 21 and 44).

Conclusions:

To summarize, we have shown that neural networks can be used to discriminate the experimental conditions under which a subject has been run and can further discriminate some subjects from each other based solely on response topography. Our initial failures to achieve either goal with RMSE data, and our subsequent success with Control Movement data, strongly suggest that the success or failure of such discrimination may be dependent upon the nature of the data. Further, our extensive use of a variety of techniques in attempting to analyze RMSE data implies that our failure was not simply due to poor choices of network architecture, parameters, or learning procedures, but is inherent in the nature of the data sets. The results of the present studies must be seen as preliminary, but constitute, we believe, a promising indication of the feasibility of the original conception.

Much more needs to be done, of course. In the short run, we would like to extend both Study 7 and Study 8 to include all subject data and all possible pairwise comparisons of subjects. We would also like to use the "Fix Procedure" described in Study 6 to determine an optimal number of hidden units to use for each kind of classification. We should then be in a position to closely compare the results of our study with the ANOVA results of the same data, classifying subjects into the three categories noted in the introduction of the paper.

In the long run, the results suggest that it may be feasible to use networks as data analytic aids to more traditional methods of analysis such as ANOVA. Further, we believe that similar techniques can be used on data derived from flight simulators, since flight simulator data channels contain records of performance that are similar in character to tracking data. In particular, we are exploring the possibility of using data from the Simulator for

Tactical Operations Research and Measurement (STORM), which has a joystick and throttle similar to that of an actual F-16D aircraft and a display system which includes an F-16 HUD (Maresh, 1992.) This system has been used to examine the effectiveness of several measures of Situation Awareness, including physiological measures (Crabtree, Marcelo, & McCoy, 1993; Stratton, Wilson, & Crabtree, 1993; Vidulich, Stratton, Crabtree, & Wilson, 1994).

One problem in analyzing flight simulator data involves the sheer amount of such data which is produced in even relatively short session times. We believe, however, that it may be possible to use neural networks to identify specific changes in pilot performance as a function of the pilot's response to external events, and to do so by monitoring the performance itself. If so, then neural networks may be able to play an expanded role in research efforts involving simulators, and may also have a role in the development of future aircraft systems that include machine monitoring of human performance.

References

Crabtree, M. S., Marcelo, R. A. Q., & McCoy, A. L. (1993). Examination of a subjective situation awareness measure during training on a tactical operations simulator. In R. S. Jensen & D. Neumeister (eds.) Proceedings of the Seventh International Symposium on Aviation Psychology. (April, 1993). Columbus, OH: Ohio State University, pp. 891-895.

Endsley, M. R. (1991). Situation awareness in an advanced strategic mission. Armstrong Laboratory, Final Report AL-T. R.-1991-0083. Wright-Patterson Air Force Base: Dayton, OH.

Maresh, J. (1992). Documentation for STORM Programs, Version 2.10. Unpublished Manuscript. [No Place Given]: Engineering Solutions, Inc.

Masters, T. (1993). Practical neural network recipes in C++. Boston: Academic Press, Inc.

Mehra, P. & Wah, B. W. (Eds.),(1992). Artificial neural networks: Concepts and theory. Los Alamitos, CA: IEEE Computer Society Press.

Netick, A. & Klapp, S. T. (1994). Hesitations in manual tracking: A single-channel limit in response programming. Journal of Experimental Psychology: Human Perception and Performance, 20, 766-782.

NeuralWare, Inc. (1993). Neural computing: A technology handbook for Professional II/Plus and NeuralWorks Explorer. Pittsburgh, PA: NeuralWare, Inc.

Rogers, S. K., Kabrisky, M., Ruck, D. W., & Tarr, G. L. (1990).An introduction to biological and artificial neural networks. Unpublished manuscript. Dayton, OH: Air Force Institute of Technology, Wright-Patterson Air Force Base.

Rumelhart, D. E. & McClelland, J. L. (1986). Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1. Cambridge, MA: MIT Press.

Stratton, M. D., Wilson, G. F., & Crabtree, M. (1993). An analysis of physiological measures in the study of pilot situational awareness. In R. S. Jensen & D. Neumeister (eds.) Proceedings of the Seventh International Symposium on Aviation Psychology. (April, 1993). Columbus, OH: Ohio State University, pp. 901-905.

Vidulich, M. A. (1989). Objective measures of workload: Should a secondary task be secondary? In R. S. Jensen (Ed.), Proceedings of the Fifth International Symposium on Aviation Psychology (Volume 2). Columbus, OH: OSU Aviation Psychology Laboratory.

Vidulich, M. A. (1990). Performance-based workload assessment: Allocation strategy and added task sensitivity. In S. Griffin (Ed.), Third annual workshop of space operations automation and robotics (SOAR '89) (NASA Conference Publication 3059). Washington, DC: National Aeronautics and Space Administration.

Vidulich, M. A. (1994). The secondary task experiment: Volume 1: Description and data formats. Unpublished manuscript. Dayton, OH: Wright-Patterson Air Force Base.

Vidulich, M. A., Stratton, B. S., Crabtree, M., & Wilson, G. (1994). Performance-based and physiological measures of situation awareness. Aviation, Space, and Environmental Medicine, May, 1994, pp. A7-A12.

Wickens, C.D. (1984). Engineering psychology and human performance. Columbus, OH: Merrill.

A MULTIPLEXED FIBER-OPTIC LASER FLUORESCENCE SPECTROMETER FOR FATE AND TRANSPORT STUDIES OF GROUNDWATER CONTAMINANTS

Brian S. Vogt
Professor
Department of Chemistry

Bob Jones University
Greenville, SC 29614-0001

Terence Tipton
Research Chemist
Environics Directorate

Bruce J. Nielsen
Environmental Engineer, Program Manager
Environics Directorate

Armstrong Laboratory
Tyndall Air Force Base, FL 32403-6001

A MULTIPLEXED FIBER-OPTIC LASER FLUORESCENCE SPECTROMETER FOR FATE AND TRANSPORT STUDIES OF GROUNDWATER CONTAMINANTS

Brian S. Vogt
Professor
Department of Chemistry

Abstract

Previous research efforts have resulted in a uniquely designed laser probe that has been interfaced to a column and used to monitor the flow of contaminant plumes through sand packed into the column. This report conveys the results of subsequent work, which has resulted in miniaturized laser probes with outside diameters of only 0.125 inch, a multiplexing system that connects several probes to the laser spectrometer, and computer control that automates instrument control, data acquisition, and data logging. Several recommendations were made that concerned further testing of the probes, direction for future experiments, and possible apparatus improvements. Additional personnel would also facilitate progress.

A MULTIPLEXED FIBER-OPTIC LASER FLUORESCENCE SPECTROMETER FOR FATE AND TRANSPORT STUDIES OF GROUNDWATER CONTAMINANTS

Brian S. Vogt

Introduction

One of the areas of ongoing research at the Environics Directorate of Armstrong Laboratory (AL/EQ) is the study of the fate and transport of contaminants in the environment. Results from the controlled contamination and monitoring of laboratory-scale model aquifers are analyzed and mathematically modeled to gain better understanding of what happens to contaminant plumes introduced into the environment through accidental spills. The monitoring of model aquifers mimics the conventional site characterization performed at solvent- or fuel-contaminated Air Force sites. At those sites wells are drilled into the ground and lined with pipe. Samples of groundwater are periodically removed from the wells and taken to laboratories for analysis. This process is both time-consuming and costly. Fiber-optic laser fluorimetry coupled with cone penetrometry promises to vastly improve site characterization so that far fewer wells can be drilled per site than are currently required.

Conventional fluorimetric analysis, or fluorimetry, is based on the excitation of aromatic organic compounds with visible or ultraviolet (UV) light followed by the emission of light referred to as fluorescence. The excitation light is typically generated by a high-pressure xenon arc lamp. Fluorimetry can be used to quantify substances because the intensity of fluorescence varies with the concentration of the emitting substance. Fiber-optic fluorimetry is a variation of conventional fluorimetry that uses optical fibers to transmit excitation light to the sample of interest. The fluorescence so generated is collected by one or more fibers and carried to optics and electronics for analysis. Unlike conventional fluorimetry, which requires that the sample be placed in a cuvette in an instrument for analysis, fiber-optic fluorimetry permits remote, continuous, on-line analysis of various substances to be performed in environments where sampling is either difficult, inconvenient, or dangerous. For example, *in situ* analysis of

44-3

biochemicals in microbial bioreactors has been performed with fiber-optic fluorimetry.[1] Lasers are used as light sources in situations where conventional source lamps are too dim to adequately excite fluorescence. The resulting technique is called fiber-optic laser fluorimetry. It has been used to demonstrate the remote sensing of polycyclic aromatic hydrocarbons (PAHs) in both atmospheric and aqueous laboratory environments.[2] Fiber-optic laser fluorimetry has been used to analyze toluene in existing monitoring wells.[3] This use of the technique eliminates the need to remove, store, and transport samples for analysis, but it does not in of itself eliminate the need for large numbers of wells.

The SCAPS (Site Characterization and Analysis Penetrometer System) program is a tri-service effort of the Air Force, Army, and Navy employing truck-mounted cone penetrometers. A cone penetrometer consists of a hollow, hardened steel probe that is hydraulically pushed into the ground at a typical rate of about one meter per minute. The penetrometer tube contains a variety of sampling and sensing devices. Optical fibers can be inserted into the penetrometer and positioned to view the soil through a small window in the side of the tube. The mobility of the truck combined with the rapid penetrometer push rate permit a fluorescence profile of the site to be obtained much more quickly and less expensively than is possible with well-based monitoring. Such systems have been successful in the field.[4,5]

Application of fiber-optic laser fluorimetry to model aquifer studies will permit rapid, noninvasive evaluation of contaminant transport. However, such experiments are time and labor intensive and must be planned carefully. It is helpful to characterize the contaminant/soil interaction with a column study before designing and initiating a model aquifer study. Work performed during the Summer of 1993 at AL/EQ resulted in a modular column system with an embedded fiber-optic laser probe.[6] After packing the column with sand, contaminant plumes containing one or two different tracer molecules were injected onto the column and pumped through with an aqueous medium. Fluorescence from the tracers was induced by UV light transmitted from a laser into the embedded probe by a communications-grade optical fiber. Multiple fibers in the probe collected the tracer fluorescence and transmitted it from the column environment to optics and electronics for analysis.

Although that system permitted relatively rapid evaluation of soil-contaminant interactions in columns, three major intrinsic limitations prevented it from being applied to model aquifer studies. First, the probe outside diameter was 0.25 inch, preventing it from fitting into the model aquifer wells in use. Second, the incorporation of only a single laser probe into the spectrometer would have required manual repositioning of the probe from well to well during experimentation. Not only would this require constant supervision, but also the repetitive insertion and removal of the laser probe would alter the flow of aqueous medium through the model aquifer. Third, instrument control, data acquisition, and data logging were all performed manually. Consequently, experiments with large amounts of data would be particularly susceptible to operator errors. Furthermore, manual control and recording limit the rate at which data can be acquired.

Subsequent research efforts have resulted in miniaturized laser probes with outside diameters of only 0.125 inch, a multiplexing system that connects several probes to the laser spectrometer, and computer control that automates instrument control, data acquisition, and data logging. This paper reports on these efforts.

Apparatus

The configuration of the apparatus, which is shown in Figure 1, is somewhat different from what was used during the Summer of 1993. Spectrometer system modifications, probe construction, probe multiplexing, computer interfacing, and the writing of custom data acquisition and control software were all performed by Dr. Terence Tipton before the Summer began. The principles of operation of the laser system has been previously described.[3] In summary, the 532 nm output of a 20 Hz frequency-doubled Nd:YAG laser (Continuum Surelite Series 20) is used to pump a dye laser containing a solution of rhodamine 590 laser dye, also known as rhodamine 6G (Exciton). The tunable, visible output from the dye laser is frequency doubled to give UV light, which is required to excite the tracer molecules under study in this laboratory. The only tracer molecule investigated in this study was 7-amino-1,3-naphthalene disulfonic acid (monopotassium salt) (Eastman Fine Chemicals, used as received), which is sometimes referred to as amino G acid (AGA). The

dye laser was tuned to give frequency-doubled output at approximately 287 nm, which is the same wavelength used in a previous study to excite both naphthalene and AGA.[6] AGA exhibits blue fluorescence and was monitored at 445 nm, which corresponds to the maximum in its fluorescence spectrum[7].
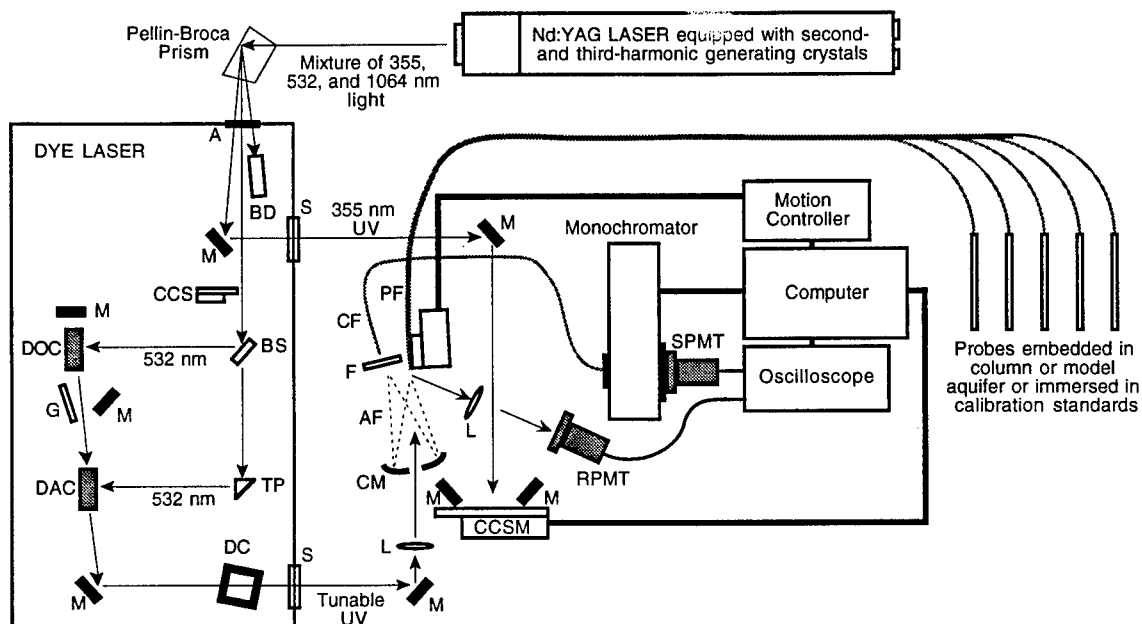
Figure 1. **Configuration of the laser spectrometer.** A = Aperture, AF = Analyte Fluorescence, BD = Beam Dump for residual 1064 nm Nd:YAG output, BS = Beam Splitter, CCS = Computer-Controlled Shutter, CCSM = Computer-Controlled Shutter and Mirror assembly, CF = Collection Fiber, CM = Collection Mirror, DAC = Dye Amplifier Cell, DC = Doubling Crystal, DOC = Dye Oscillator Cell, F = Filter, G = Grating, L = Lens, M = Mirror, PF = Probe Fibers mounted on motorized turret interfaced to motion controller, RPMT = Reference Photomultiplier Tube, S = Shutter, SPMT = Sample Photomultiplier Tube, TP = Turning Prism.

The UV output from the dye laser was focused onto the launch end (excitation end) of a probe made with fiber of core diameter of 600 μm and numerical aperture of 0.4 (Fiberguide Industries). One of the probes used in previous studies had an outside diameter of 0.5 inch and contained two fibers (one for excitation and one for collection).[8] Another was 0.25 inch in diameter and had a bundle of seven fibers, two of which were used for excitation and five for collection.[6] That design was modified to give the current design (illustrated in Figure 2), which is only 0.125 inch in diameter and has a single fiber for both excitation and collection.

Figure 2A shows how the fiber is clamped to the body of the probe with a compression fitting. This permitted quick disassembly and reassembly if either the fiber or the probe end was damaged. As seen in Figure 2B, the fiber is centered in the body of the probe with a cylinder of epoxy. Also illustrated in that Figure is the cone of excitation light from the fiber. It just fills the glass mirror, which is placed approximately 8.5 mm from the bottom of the sapphire window. Positioning the mirror in this way prevents any excitation light from traveling around the mirror and reflecting unpredictably off the walls of the sample container, which varies from glass

Optical fiber

Knurled end nut

Stainless steel ferrule

Rubber o-ring

Cajon ultra-torr stainless steel fitting for 1/8" O.D. tubing

Stainless steel probe body, 10 inches long (1/8" O.D. tube)

Optical fiber positioned inside the probe body

Probe body

Probe body ground away to leave only a thin strip of metal

Mirror mount formed by grinding the probe body down

Plug of epoxy formed around the fiber and ground to tight tolerance to prevent the fiber from wobbling inside the probe body

Sapphire window glued over end of tubing to prevent liquid from getting inside the probe body (shown removed for clarity)

Cone of excitation light from the fiber

Mirror made of glass microscope slide

Chip of fluorescent glass from a lens of a pair of laser goggles

(A) Overview.   **(B) Details of the Tip.**
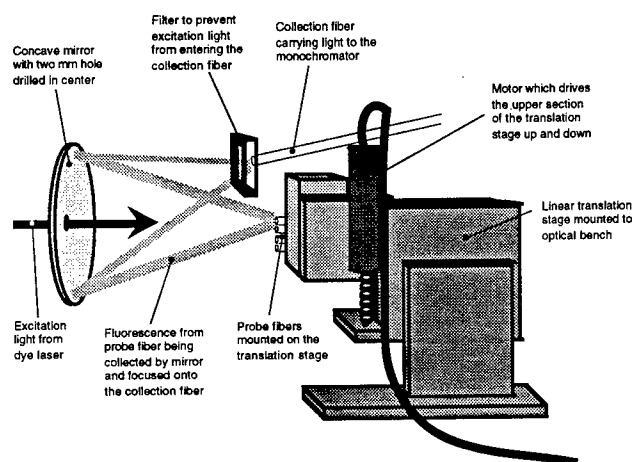**Figure 2. Design of the laser probe.**

volumetric flasks containing calibration standards to screen-mesh cylinders designed to prevent the optical path from being filled with sand. The fiber thus has a constant, controlled optical view of fixed pathlength. One shortcoming of this design is that it cuts sensitivity by decreasing the volume of solution irradiated by the excitation light. However, there is a

positive side to this as well. As the concentration of the analyte increases, some of its fluorescence is reabsorbed by the solution instead of reaching the fiber to be carried back to the detector. This phenomenon (known as the inner-filter effect) effectively limits the depth of solution from which the fiber can collect fluorescence. The spatial resolution (depth perception) of a single-fiber probe varies with analyte concentration because the magnitude of the inner-filter effect varies with concentration.[9] Limiting the depth of view of the fiber artificially by introducing the mirror in its view provides an essentially fixed spatial resolution of about 8.5 mm (the approximate distance from the fiber tip to the center of the mirror) until the concentration of the analyte becomes very high, which is not anticipated in our studies. Although the shorter pathlength does cut sensitivity somewhat, it is advantageous to have good spatial resolution in column and model aquifer studies so that pseudo-point concentrations can be obtained instead of those averaged over long distances.

The chip of fluorescent glass illustrated below the glass mirror in Figure 2B is present to give a reference fluorescence signal to be used in making corrections for light scattering in turbid solutions. The principle and practice of making such corrections has been previously described.[6] Although time prevented using this feature of the probe, several comments are in order. The previous use of this design relied on rhodamine 590 dye embedded in epoxy behind the glass mirror instead of a chip of fluorescent glass. Photolysis of rhodamine 590 caused a degradation in the reference signal. This has also been observed in thin sol-gel films doped with rhodamine 590.[10] It is hoped that the fluorescent glass will provide a more stable reference signal. The 355 nm third-harmonic from the Nd:YAG (see Figure 1) was generated to excite the chip of fluorescent glass. However, it was blocked by a shutter to prevent such excitation. Such excitation by the 287 nm light from the dye laser was prevented by the glass mirror on the probe, which effectively reflects and/or absorbs that wavelength. Consequently, there was no detectable emission signal from the chip of fluorescent glass throughout this study. The computer-controlled shutter and mirror assembly, illustrated in Figure 1, was present so that the 355 nm light could be directed to the excitation end of the probe while simultaneously

blocking the dye laser light during turbidity measurements. This part of the apparatus was neither tested nor used in this study. The design of the probe-spectrometer interface is shown in Figure 3. Excitation light from the dye laser is focused onto the launch end of a probe fiber after passing through a 2 mm diameter hole drilled in a mirror 50 mm in diameter. This is similar to designs used elsewhere.[11,12] The excitation light passes through the fiber to the probe and into the solution. The fluorescence so generated is scattered isotropically into the solution. That portion of the fluorescence that is scattered back into the fiber travels back to the optical bench, where it is collected and focused onto the collection fiber. Only a small percentage of the fluorescence is lost through the mirror hole. A filter prevents any back-scattered excitation light from entering the collection fiber.

Figure 3. Probe-Spectrometer Interface.

The fluorescence is carried by that fiber to a computer-controlled monochromator (Chromex 500 IS/SM) for dispersion into its component wavelengths, detected by the sample photomultiplier tube (PMT), and displayed on a digital storage oscilloscope (Tektronix 2440) (see Figure 1). As illustrated in Figure 3, multiplexing was accomplished by epoxying a group of probe fibers together in a linear array and mounting it vertically on a linear translation stage. The stage was moved up and down to position a given fiber in the optical path by a motorizer (Newport 860-1) driven with a two-axis motion controller (Newport 860-C2) interfaced to a computer (generic 486 PC) through a parallel port. Custom software written in BASIC (Microsoft) was used to drive the motion controller. Translation stage position information was obtained by a calibration procedure. The calibration was performed in one of two ways, depending upon the experiment. Launch-end calibration involved

monitoring excitation light reflected off the surface of the fiber faces to the reference PMT (Figure 1), which was equipped with a UV filter to prevent ambient visible light from interfering. The efficiency of reflection of the excitation light from the surface of the fibers is higher than from the fiber cladding, fiber buffer, and epoxy between the fibers. Consequently, the signal at the reference PMT rises and falls as the fiber array is moved through the path of the excitation light. Each peak corresponds to a different fiber. The reference position for time zero is a hardware limit on the motorizer. Positioning a given fiber is accomplished by instructing the motion controller to drive the motorizer for the period of time required to reach the appropriate maximum in the calibration data file. Probe-end calibration is identical to launch-end calibration except that the fluorescence from a solution of AGA was monitored with the probes and the sample PMT instead of looking at reflections off the launch surfaces with the reference PMT.

Several different types of interfaces were used to integrate the computer into the apparatus. Control of the oscilloscope and extraction of data from it were achieved through a GPIB card (National Instruments GPIB-PCII/IIA). The Nd:YAG power supply, computer-controlled shutter in the dye laser, and computer-controlled shutter and mirror assembly (see Figure 1) were controlled *via* an I/O relay board (Industrial Computer Source DIO8-P). The dye laser shutter consisted of a modified control valve. The monochromator was interfaced *via* a serial port. All of these components were controlled with the same BASIC program used for translation stage positioning.

The control software was used to set or select many experimental parameters. Oscilloscope gain was set automatically by reading several traces and adjusting the gain if the signal was too high or low. Signal averaging was performed with the control software during gain checking and by the oscilloscope itself once the gain was properly set. The fluorescence signals were quantified by measuring the voltage at the maximum of the decay profile on the oscilloscope. This was done reproducibly by using the oscilloscope-displayed profile of excitation light reflected off the launch end of the fiber (detected by the reference PMT) as a time reference. In this study all fluorescence measurements were made on the AGA decay profile at a time 48 ns

past the half-maximum of the leading edge of the reference profile. Monochromator wavelength and slit widths were also set through the software.

Plots of calibration data or column data were automatically generated and displayed on the computer monitor as the data were collected. This alerted the operator to problems so that the measurements could repeated (if the problems could be rectified) or aborted (if the problems could not be fixed). Data were automatically logged by the computer after each measurement. A separate data file was created for each probe and was opened and closed as data were collected to minimize the likelihood of losing data during power interruptions, which are not uncommon in the summer at AL/EQ.

Because fluorescence intensity varies with excitation power, it was necessary to monitor the laser power and normalize the fluorescence signals to correct for any variation in laser power. This was done by recording the magnitude of the signal at the reference PMT whenever a reading was taken from the signal PMT.

Details of the stainless-steel column and pump apparatus and the probe-column interface can be obtained in the full report filed on base.[13] The background solution consisted of distilled water containing 0.005 M $CaSO_4$ (calcium sulfate) and 0.02 w/w % NaN_3 (sodium azide). The $CaSO_4$ serves to raise the ionic strength of the solution; the NaN_3 was added to prevent microbial and algal growth. This mixture has been used successfully in batch, column, and model aquifer sorption studies.[14] All calibration standards and contaminant plumes were made up in background solution.

Results

Normalization of the fluorescence signals to compensate for fluctuations in laser power was found to be successful except when laser power dropped significantly. This occurred when the temperature in the laboratory rose throughout the day due to poor air conditioning. Our measurements indicated that the Nd:YAG power did not change much. Consequently, the apparent drop was related to the dye laser. The UV output of the dye laser is very sensitive to the tuning angle of the doubling crystal (see Figure 1), which is affected by ambient temperature. Several options could be pursued. It is possible to monitor the laser power more accurately by using an evanescent

wave coupler to monitor laser power near the sample end of the probe.[15] This, however, would increase both probe size and complexity, neither of which are acceptable in this situation. Another possibility would be to integrate the doubling crystal into a computer-controlled stage to permit automated tuning. This option is going to be pursued.

We have found several drawbacks to the modular probe design. Although easy disassembly and reassembly facilitates repair, it also presents more opportunities for the probe end of the fiber to be damaged. Probe performance is very sensitive to the condition of the fiber. AGA calibration data obtained before small scratches on the end of one of the probes were removed were noisy and had a low slope; after remove the scratches the noise was gone and the slope was significantly higher. Compression of the fiber by the o-ring holding it in place can also damage the fibers and degrade probe performance. On fiber this resulted in twisting and tearing to the fiber cladding and buffer. The damaged end of the fiber was cut off just above the damage so that it could be reinserted into the probe body. Also visible were scratches to the cladding and buffer, which could have been caused by the fiber rubbing on the inside edge of the stainless-steel probe body during assembly or disassembly. In one or two instances it was found that dirt worked down into the probe body and interfered with the passage of light in and out of the probe. Sometimes the fibers were not mounted in the probes tightly, in which case movement of the fiber tip away from the sapphire window and/or rotation of the fiber around the long axis of the probe changed probe response. Fiber rotation is a problem because it is difficult to exactly center the end of the probe fiber in the epoxy cylinder (see Figure 2B). Some of these problems were solved by applying glue to the outside edge of the fiber and attaching it to the inside of the sapphire window. A small length of heat shrink was carefully applied to the outside of the fiber to protect it where the o-ring clamps it in place. Although these changes improved probe reliability, they make it impossible to quickly disassemble and reassemble the probes if they are accidentally damaged. Furthermore, some instability in signal was still observed. The fluorescence from a standard solution of AGA was found to fluctuate with time when monitored with any given probe. It may be that the fiber core is pistoning inside the cladding. This problem remains

unresolved but may be fixed by devising a better method of holding the fiber inside the probe. It was also found that the windows and mirrors came off relatively easily. Previous designs did not suffer from this, probably because the larger diameter provided more surface area for gluing.

Initially, the precision of the translator in positioning of the fibers in the path of the excitation light was found to be unacceptable. Variation in the position of any given fiber changes the amount of excitation light entering the probe, which directly changes the intensity of fluorescence generated at the end of that probe. Furthermore, the amount of fluorescence collected by the perforated mirror and focused onto the collection fiber is also changed when the fiber is mispositioned. These problems are translated into the signal changing after fiber repositioning and is illustrated in Figure 4. This graph shows the fluorescence from a 459 ppb AGA solution as a function of time for five probes. The fiber from probe 1 was the first fiber in the bundle and as such was closest to the motorizer hardware limit used as a time reference; probe 5 was farthest away from the time reference. The increase in fluctuation in going from probe 1 to probe 5 is explained by assuming that error in positioning from fiber to fiber is additive. It was found that, due to a software problem, the motorizer was not being moved completely to its hardware limit. The software was modified to remedy this.

Comparison was made between the two translator calibration procedures. Probe-end translator calibration resulted in more stable fluorescence signals than did launch-end translator
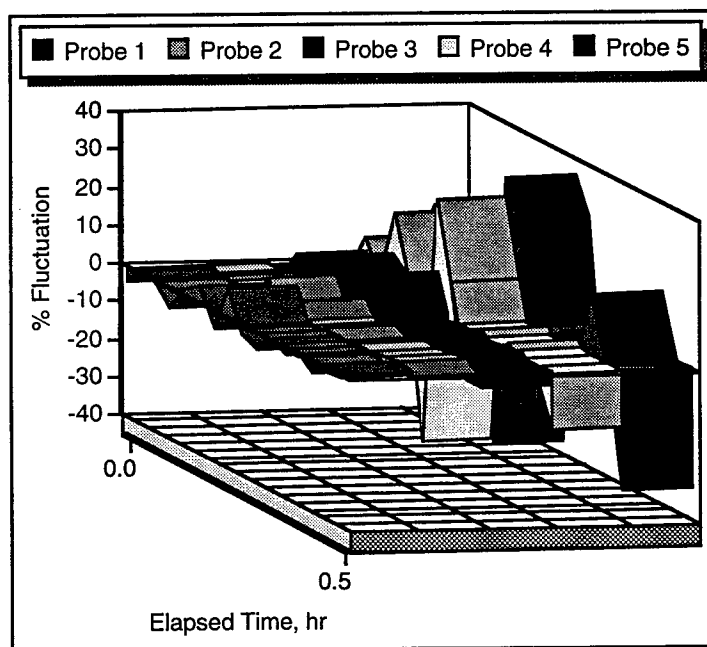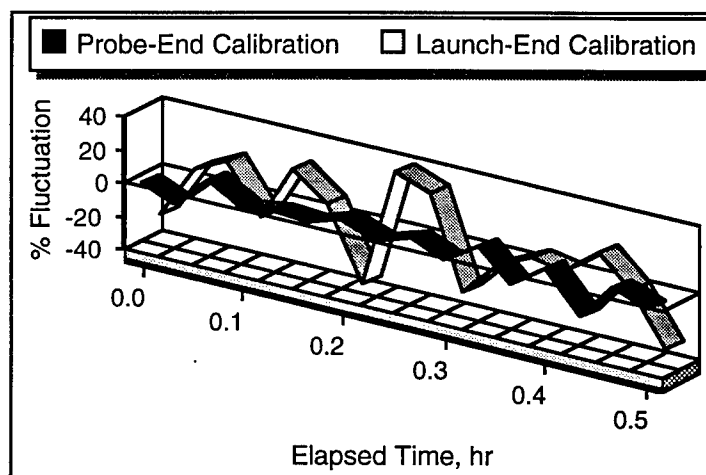

Figure 4. **Effects of translator imprecision on signal.**

calibration. This is illustrated for probe 5 (the worst case probe in Figure 4) in Figure 5. Launch-end calibration assumes that the highest degree of reproducibility in fiber positioning is obtained when the excitation light enters the center of the fiber. Probe-end calibration assumes that positioning precision is maximized when the fluorescence signal from the distal end of the probe is maximized. Not evident in Figure 5 is the fact that the signal from the 459 ppb AGA solution with launch-end calibration was only half that with probe-end calibration, which is consistent with the assumptions. Furthermore, the magnitude of the variation in these data clearly show that probe-end calibration is superior. This suggests that the probes are not homogeneous along the long axis of the fiber, resulting in higher light transmission and correspondingly higher signals when the light is launched somewhat off the center of the fiber end.

Figure 6 shows signal stability data after both translator calibration software enhancements were made and probe-end translator calibration was phased in. The probe 4 data were aberrant, suggesting that something was wrong with that probe. Consequently, those data were not included in Figure 6. It is noteworthy that the data for probe 6, which was furthest from the motorizer time reference point, are significantly better than were the data for probes 4 and 5 before the improvements were made.

Figure 5.   Signal stability with different translator calibrations.

We had various other hardware problems. The Nd:YAG laser underwent auto-shutdown on several occasions due to decreased flow in its internal cooling loop. Cleaning an internal filter screen improved but did not completely remedy the situation. The translator motorizer failed once due to the breakage of a
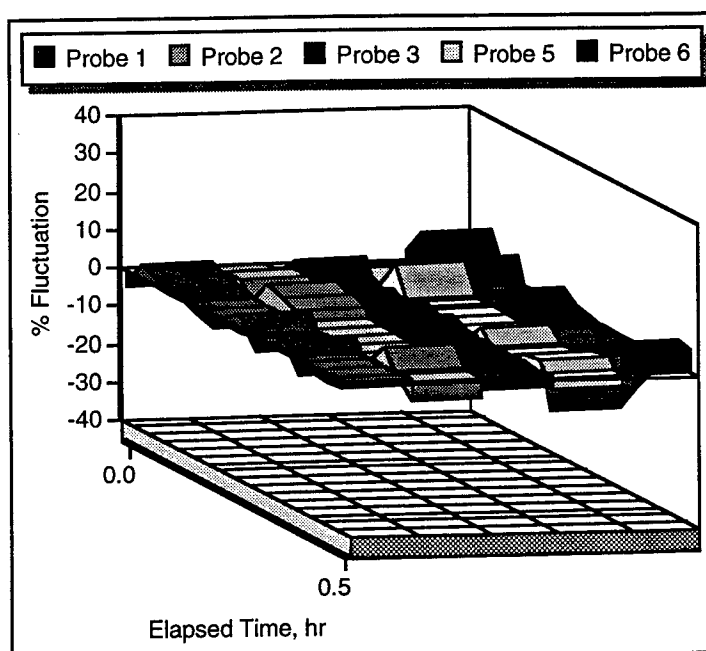
shaft pin. This was repaired, For no apparent reason the monochromator locks up sometimes and does not respond to the computer, requiring a monochromator restart. Sometimes the computer locks up due to an electrical problem related to the dye laser shutter. On several occasions the 3 1/2 inch floppy drive on the computer would not respond. This requires a computer reboot and one time resulted in the loss of some data files. All of these problems remain unresolved except that of the motorizer.

Limits of detection for four probes were estimated from typical AGA calibration data using the IUPAC method[16] and are 4.1 ppb (probe 1), 4.9 ppb (probe 2), 7.6 ppb (probe 4), and 1.1 ppb (probe 6). Although there were five ports in the column, only four probes were embedded because one of the ports had excess solder on the inside of the mesh screen cylinder, making it virtually impossible to insert a probe into it without bending the steel support holding the glass mirror in position. The limit of detection estimated in 1993 using a probe with an outside diameter of 0.25 inch was 0.17 ppb. The significantly poorer detection limits obtained with the 0.125 inch diameter probes are attributed to the shorter path length, which was necessary for reasons already discussed.

Figure 6. Signal stability after translation stage calibration improvements.

Figure 7 shows breakthrough curves for AGA on a sand column. The pumping rate was 0.2 mL/min. The contaminant plume was a 2.895 mL pulse of 1530 ppb AGA. These data are qualitatively good. The plume was the narrowest as it passed probe 1 because it was the first probe in the column. Broadening
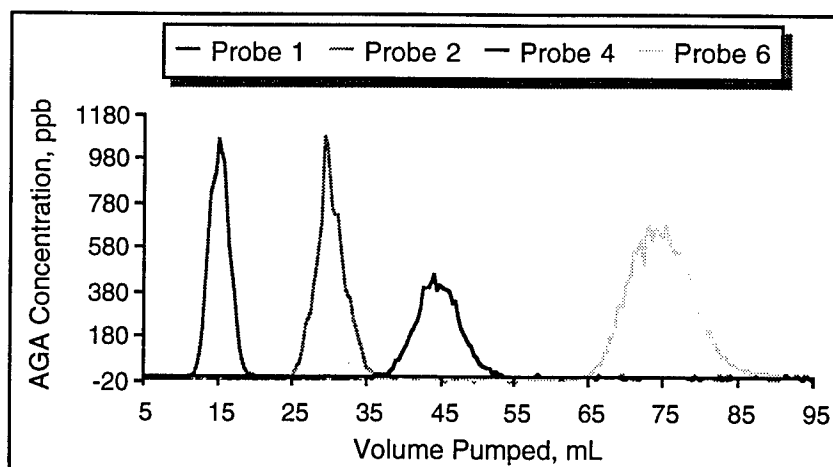
Figure 7. Breakthrough of AGA on sand in a column with four embedded probes.

caused by diffusion of the AGA in the liquid phase occurred as the plume passed through the column. The spacing between the peak maxima is constant because of the uniform spacing of the probes in the column. The gap between the maxima observed by probes 4 and 6 is twice the other gaps because of the empty port previously referred to. It should be pointed out that a power outage occurred near the beginning of this run and the background pump was not running for a period of 15 minutes. This has been compensated for in the calculations so that values of the volume pumped shown on the x axis are correct. However, these profiles are probably slightly broader than they would have been in the absence of the power outage.

Trapezoidal integration was used to estimate the mass of AGA represented by each breakthrough curve. Percent recovery was then calculated by comparing the mass estimated by integration to the mass of AGA known to be injected into the column. Three of the four percent recoveries were either unreasonably high or low. It was decided to recalibrate the probes twice to see if any other variables were affecting probe performance. The calibration performed just prior to the column run was done by lowering probes oriented vertically into volumetric flasks containing AGA standards. The second calibration, performed two days after the column run, was performed with the volumetric flasks clamped almost horizontally and the probes inserted into the flasks from the side. This calibration simulated the position of the probes in the column and will be referred to as the semihorizontal calibration. The

screened calibration was performed immediately after the semihorizontal calibration and was done with the probes oriented vertically but with their ends covered with a screen mesh cylinder to simulate the column environment. The column data were then reanalyzed using the new calibration data and percent recoveries recalculated. The results are shown in Figure 8.

	Probe Configuration During Calibration		
	Vertical	*Semihorizontal*	*Screened*
Probe 1	80.3%	72.8%	83.3%
Probe 2	95.3%	94.0%	96.3%
Probe 4	70.8%	62.8%	75.7%
Probe 6	160.3%	139.5%	139.4%

Figure 8. Column recoveries obtained with different calibration methods.

Probe 2 is the only probe that gave stable calibration behavior (identical calibration curves with all three methods). Therefore, it is not surprising that the three column recoveries estimated for probe 2 are close to 100% and essentially identical. For the other probes all three calibrations differed from each other significantly. One cannot explain this on the basis of a variation in instrument performance because there is no clear trend in how they differed from each other. As already suggested, it may be that the fiber core is pistoning inside the cladding. Different degrees of core pistoning combined with small variations in light losses (5-10%) due to stress on the fiber at the point of o-ring compression could explain most of the variations observed. Probe 6 merits additional comments. Computer-displayed plots of the breakthrough data observed during the column run made it clear that laser power was dropping significantly as the run progressed. It was decided to tweak the dye laser doubling crystal to see whether power would increase significantly and also to ascertain whether the signal normalization procedure would work if laser power did increase significantly. Crystal tweaking was done just before the plume began to reach probe 6. It was observed that the laser power increased by roughly 250% to near what it was during the three calibrations. Although one would the signal normalization procedure to work especially well ·under these conditions, the fact that the recoveries came out so high means that it did not. The fact that the

recoveries estimated with all three calibrations were excessively high suggests that the fluorescence signals from this probe during the column run were high. Perhaps this is due to a change in laser beam spatial characteristics.

Conclusion

Overall, the probe multiplexing and automated data acquisition and logging were judged to be successful. Although several problems related to the computer and attached hardware remain, it is expected that they can be solved. Work is underway to improve the probe design by using silica-clad fiber. Although it seems that this fiber is somewhat more brittle than plastic-clad fiber of the same core diameter, it is hoped that the cladding is so tightly adhered to the core that core pistoning will be eliminated. It is also hoped that the silica cladding will attenuate the UV excitation light less than plastic cladding. Further improvements may also make the probes less susceptible to fiber damage.

Several courses of action are recommended. First, the miniaturized probe design has yet to be quantitatively evaluated in its ability to correct for light scattering due to the presence of suspended solids in turbid samples, which may be encountered during model aquifer studies. Experiments similar to those performed with the 0.25 inch diameter probe could be done.[6] Second, the response of the miniaturized probes to naphthalene needs to be evaluated. It is used in column and model aquifer studies because it is an important component of jet fuel and is often found at contaminated Air Force sites. Third, perform a model aquifer study with a contaminant plume containing both AGA and naphthalene. Conventional monitoring using gas chromatography (GC) or another traditional method of analysis needs to be used to validate the use of the fiber-optic laser spectrometer in such studies. Fourth, consider an alternate multiplexing scheme if higher reliability and faster data acquisition are needed. Some researchers have placed bundles of fibers in the paths of light beams large enough to cover several fiber faces at once. One group was using fibers with core diameters of 250 μm for near infrared (NIR) analysis of sugar in fruit using a tungsten-halogen lamp as a spectroscopic source.[17] Another group used 50 μm diameter fibers and

illuminated the faces of the fibers with laser light directed through a microscope objective.[18] Both groups used charge-coupled device (CCD) cameras to image light from all of the fibers at once. Our research requires larger diameter fibers to achieve sufficiently high light throughput with UV excitation. Consequently, a high-power large-diameter laser beam would be required to enable the simultaneous illumination of all the fibers in a bundle of larger (600 μm) diameter fibers. This is not available from our dye laser. If it is possible to simultaneously excite both naphthalene and AGA with a 266 nm light, then an excitation beam of sufficient power and diameter could probably be obtained by using the frequency-quadrupled output of the Nd:YAG laser. A solid-state, tunable UV laser with a beam diameter of 2-3 mm[19] might also be an adequate source for a study with 10-15 fibers. A CCD camera is already on hand in our laboratory. One drawback of this type of multiplexing scheme is that a CCD camera is somewhat less sensitive than a PMT. A sufficiently powerful laser could probably compensate for this to some degree because it can excite more fluorescence than can a lower power laser. Furthermore, multiplexing with a CCD camera in this way would completely eliminate the need for a translation stage and probably improve signal stability quite significantly, which would improve sensitivity during column and box runs. Fifth, progress would be facilitated if more personnel were available to assist in this effort.

References

1. J.J. Horvath and C.J. Spangler, "In Situ On-line Optical Fiber Sensor for Fluorescence Monitoring in Bioreactors," in **Frontiers in Bioprocessing II,** P. Todd, S.K. Sikdar, and M. Bier, Editors, American Chemical Society (ACS) Conference Proceedings Series, 99-115 (1992).
2. U. Panne, F. Lewitzka, and R. Niessner, "Fibre Optical Sensors for Detection of Atmospheric and Hydrospheric Polycyclic Aromatic Hydrocarbons," **Analusis, 20,** 533-542 (1992).
3. G.D. Gillispie and R.W. St. Germain, "In-Situ Tunable Laser Fluorescence Analysis of Hydrocarbons," **Environmental and Process Monitoring Technologies,** T. Vo-Dinh, Editor, **Proc. SPIE 1637,** 151-162 (1992).
4. S.H. Lieberman, G.A. Theriault, S.S. Cooper, P.G. Malone, R.S. Olsen, and P.W. Lurk, "Rapid, Subsurface, In Situ Field Screening of Petroleum Hydrocarbon Contamination Using Laser Induced Fluorescence Over Optical Fibers," **Proceedings of the Second International Symposium on Field Screening Methods for Hazardous Wastes and Toxic Chemicals,** 57-63 (1991).
5. W.L. Bratton, J.D. Shinn, S.M. Timian, G.D. Gillispie, and R. St.

Germain, "The Air Force Site Characterization and Analysis Penetrometer System (AFSCAPS); Laser Induced Fluorescence Cone Penetrometer. Volume I-System Development and Evaluation," Air Force technical report AL/EQ-TR-1993-0009 (1993).

6. B.S. Vogt, "Soil Column Studies With A Fiber-Optic Laser Spectrometer," *USAF Summer Research Program--1993, Summer Research Program Final Reports, 2, Armstrong Laboratory,* 2-24 to 2-44. This paper will also be available as Air Force technical report AL-EQ-TR-1994-0011 (1994) (in press).

7. P.L. Smart and I.M.S. Laidlaw, "An Evaluation of Some Fluorescent Dyes for Water Tracing," *Water Resources Research, 13,* 15-33 (1977).

8. B.S. Vogt, "Application of Fiber-Optic Laser Fluorescence Spectroscopy to Environmental Monitoring," *USAF Summer Research Program--1992, Summer Faculty Research Program (SFRP) Reports, 6,* Air Force technical report AFOSR-TR-93-0116 (1992) (NTIS accession number AD-A261 994/8/XAB).

9. Z.Y. Zhu and M.C. Yappert, "Determination of Effective Depth and Equivalent Pathlength for a Single-Fiber Fluorometric Sensor," *Applied Spectroscopy, 46,* 912-918 (1992).

10. U. Narang, F.V. Bright, and P.N. Prasad, "Characterization of Rhodamine 6G-Doped Thin Sol-Gel Films," *Applied Spectroscopy, 47,* 229-234 (1993).

11. J. Louch and J.D. Ingle, Jr., "Experimental Comparison of Single- and Double-Fiber Configurations for Remote Fiber-Optic Fluorescence Sensing," *Analytical Chemistry, 60,* 2537-2540 (1988).

12. J.R. Bowyer, J.P. Alarie, M.J. Sepaniak, T. Vo-Dinh, and R.Q. Thompson, "Construction and Evaluation of a Regenerable Fluoroimmunochemical-based Fibre Optic Biosensor," *Analyst, 116,* 117-122 (1991).

13. B.S. Vogt, T. Tipton, and B.J. Nielsen, "A Multiplexed Fiber-Optic Laser Fluorescence Spectrometer for Fate and Transport Studies of Groundwater Contaminants." For a copy of this full-length report, contact Dr. Terence Tipton at AL/EQC, 139 Barnes Drive, Suite 2, Tyndall AFB, FL 32403-5323 PO Stop 37.

14. W.G. MacIntyre, T.B. Stauffer, and C.P. Antworth, "A Comparison of Sorption Coefficients Determined by Batch, Column, and Box Methods on a Low Organic Carbon Aquifer Material," *Ground Water, 29,* 908-913 (1991).

15. H. Xu and J.E. Kenny, "Measurement of Incident Intensity of UV Radiation Delivered by Optical Fibers," *Applied Spectroscopy, 45,* 429-434 (1991).

16. J.D. Winefordner and G.L. Long, "Limit of Detection: A Closer Look at the IUPAC Definition," *Analytical Chemistry, 55,* 712A-724A (1983).

17. V. Bellon, J.L. Vigneau, and M. Leclercq, "Feasibility and Performances of a New, Multiplexed, Fast and Low-Cost Fiber-Optic NIR Spectrometer for the On-Line Measurement of Sugar in Fruits," *Applied Spectroscopy, 47,* 1079-1083 (1993).

18. J.A. Taylor and E.A. Yeung, "Multiplexed Fluorescence Detector for Capillary Electrophoresis Using Axial Optical Fiber Illumination," *Analytical Chemistry, 65,* 956-960 (1993).

19. T. Meguro, T. Caughey, L. Wolf, and Y. Aoyagi, "Solid-State Tunable Deep-Ultraviolet Laser System from 198 to 300 nm," *Optics Letters, 19,* 102-104 (1994).

INVESTIGATION OF THE ROLE OF HAPTIC MOVEMENT
IN TACTILE PATTERN PERCEPTION

Janet M. Weisenberger
Associate Professor
Department of Speech and Hearing Science

Ohio State University
1070 Carmack Road
Columbus, OH 43210

INVESTIGATION OF THE ROLE OF HAPTIC MOVEMENT
IN TACTILE PATTERN PERCEPTION

Janet M. Weisenberger
Associate Professor
Department of Speech and Hearing Science
Ohio State University

Abstract

A series of experiments was conducted to investigate the role of haptic movement in the perception of complex vibratory patterns. In initial studies, results indicated that haptic movement of the display relative to the pattern to be sensed was superior to passive movement of the pattern on a stationary fingertip. In the present experiments, two further issues were addressed. First, we attempted to determine what aspects of haptic scanning might have produced superior performance. Possible candidates included the ability to repeat the stimulus under haptic scanning, and the ability to scan the stimulus from any direction, two aspects not available under the other modes. Two experiments addressed these aspects of presentation. In an experiment in which observers were permitted to repeat the stimulus under any presentation mode, results indicated that the ability to have multiple presentations of the stimulus accounted for much of the benefit provided by the haptic mode. A second experiment, in which subjects were permitted to select the direction of scanning in a passive scanning mode, revealed that certain scanning directions produced higher levels of performance, suggesting that they carried more information about stimulus features.

In a second series of studies, the question of whether the tactile "field of view" could be reduced was investigated. Reductions of the original 30-element tactile display, which covered the entire fingertip, to 9 elements, 4 elements, and 1 element were implemented, and pattern perception under haptic scanning was evaluated. Performance was unimpaired for reductions to 9 elements, and only slightly impeded by reductions to 4 elements. However, scanning with a single element was very difficult, and performance levels barely exceeded chance. Results are discussed in terms of practical development of tactile displays for telerobotics and virtual reality applications.

INVESTIGATION OF THE ROLE OF HAPTIC MOVEMENT
IN TACTILE PATTERN PERCEPTION

Janet M. Weisenberger

Under most circumstances, normal human sensing of object properties of shape and surface texture is accomplished via active movement of the skin surface (typically the fingertips) along the object to be sensed. This motion appears to be essential for accurate sensing of object properties.

Nonetheless, a considerable literature has arisen from studies of the tactile system's ability to resolve stimuli presented in a passive mode to a stationary skin surface. The motivation for this line of research has been twofold. One driving force has been the desire to measure the spatial, temporal, and intensitive data-handling characteristics of the tactile system under stimulus conditions analogous to those employed in studies of the visual and auditory modalities. Another impetus has been the development of devices to use the tactile system as an alternative communication channel for persons with auditory or visual impairments. Such devices, which transmit speech or printed information via spatiotemporal tactile patterns to substitute for the impaired modality, are typically placed on the fingertip or other sensitive body area. Perception of the patterns presented on these devices does not require active movement by the observer.

Much of the work in identification of complex tactile patterns presented on stationary displays has been performed by Craig and his colleagues (Craig, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1985; Craig & Evans, 1987; Evans & Craig, 1986). Craig's work has employed the two-dimensional vibratory display of the Optacon (Telesensory Systems), a reading aid for blind persons. The display consists of 144 piezoelectric stimulators, vibrating at a frequency of 230 Hz, that contact the index fingertip of the user. In studies of pattern recognition with this display, Craig has mapped the time course and spatial extent of recognition masking and examined attentional determinants of tactile processing. Although the finger is placed on a stationary display, movement of tactile patterns can be simulated by spatiotemporal pattern variation (e.g., "scanning" or "drawing" movements).

Perhaps one of the most surprising findings of these pattern recognition studies, in light of the abovementioned importance of movement between skin and stimulus, was Craig's (1981) report that pattern identification accuracy was better for a "static" stimulus (no simulated movement) than for any of four presentation modes that employed simulated movement: scan, slit-scan, continuous-sequential, and discontinuous-sequential drawing modes. It might have been expected that the introduction of a movement cue would bring the stimulus situation closer to that typically encountered in the real world, and thus that accuracy would be enhanced.

In Craig's 1981 study, stimuli were presented over a range of durations from 4 to 1000 ms. At long durations (greater than 400 ms), most of the modes of presentation showed similar levels of performance.

45 - 3

As duration decreased, however, performance differences emerged. The static mode was relatively unaffected by duration down to 25 msec, below which performance levels dropped markedly. In the scan mode, reducing duration below 400 ms resulted in a systematic, and substantial, drop in performance. Performance in the slit-scan, continuous-sequential, and discontinuous-sequential modes also decreased at shorter durations.

The results were in contrast to earlier findings of Loomis (1974), who presented patterns on the tactile display of the TVSS, a visual substitution device for blind persons that converted optical stimuli to patterns of vibration on a 20 x 20 display worn on the back (White, Saunders, Scadden, Bach-y-Rita, & Collins, 1970). Four modes of presentation were employed: static, scan, and two slit-scan modes. Loomis found that the static and passive scan modes yielded much worse performance than the two slit-scan modes, and argued that the tactile system performed poorly when forced to rely only on spatial frequency information for pattern identification. Loomis proposed that the two slit-scan modes provided phase information in addition to spatial information, and thus produced better performance. Earlier results from Beauchamp, Matheson, and Scadden (1971) supported this notion. However, in Craig's 1981 study, the slit-scan mode actually yielded the worst performance.

Two possible differences might account for the results in the two studies. First, data from either study may have been device-specific. Second, Loomis' stimuli were much longer in duration, exceeding 1 s. At these durations Craig found static and scan modes to be comparable, although the slit-scan mode still produced poorer performance even at this duration. An additional complication is the report of Loomis (1980), who used the Optacon display in a study of pattern perception with stimuli of different sizes, and found that small patterns were more recognizable when the slit-scan presentation mode was used, but that the static mode produced better performance for larger letters. In this study, durations were again long (1.5 s) relative to Craig's durations. However, the use of the Optacon display in Loomis' 1980 study suggests that the results are not specific to a particular device.

Another possibility to consider is that the simulated movement cues may not have been good approximations of active, or haptic movement by the human observer. Sherrick and Rogers (1966) and Kirman (1974) reported interstimulus element intervals of 50-100 ms for optimal perception of movement. In Craig's moving stimulus presentation modes the intervals between successive stimulus elements were much shorter at brief total stimulus durations. It is possible that the stimulus perceived with real haptic movement is quite different from that presented under a passive scan condition. Nonetheless, it would seem that active control of the movement would provide additional kinesthetic or motor cues to facilitate perception.

The relative effects of active and simulated stimulus movement on identification of tactile patterns were investigated by Weisenberger and Hasser (1994) using a tactile array designed for telerobotic and virtual environment applications. This display differed from that of the Optacon used by Craig and Loomis

in several respects. First, the Optacon stimulators were piezoelectric bimorphs with a very narrow stimulus bandwidth, whereas the stimulators in the display used in the present study were a titanium-nickel shape-memory alloy (SMA) with a broader bandwidth. These were activated by superimposing a pulse-width modulated current on an initial DC. Unlike the Optacon's fixed-frequency 230 Hz signal, frequency and duty cycle of the SMA display's stimulus, as well as the duration of the initial DC component, can be varied in software. This flexibility permits evaluation of pattern presentation conditions in which the SA mechanoreceptors might play a dominant role (i.e, by using low vibration frequencies).

The SMA display has been mounted on the pointer of a digitizing pad, and "virtual" patterns can be constructed in software such that active scanning movements of the display across the digitizing pad can encounter a pattern and move across it in any direction. This aspect of the apparatus permits the inclusion of an active, or haptic, movement presentation mode in the pattern identification task. Identification of a set of pattern stimuli was compared for three modes of presentation: static, passive scan, and haptic scan. Vibration frequency for the pattern stimuli was varied, to provide conditions in which all three tactile mechanoreceptor populations (SA, RA, PC) were activated. Results indicated that the haptic mode produced superior performance when a set of relatively complex patterns (letters of the alphabet) was employed. Weisenberger and Hasser cited a number of possible factors that may have contributed to the superiority of the haptic mode. These included the fact that under the haptic mode, observers were unconstrained in their behavior, and could scan the stimulus pattern as often as desired before responding. Further, they could scan the pattern in any direction they wished. In addition, total stimulus duration, given the repeated scanning, was longer for the haptic mode.

One aim of the experiments performed in the present study was to address some of these factors to determine the basis for the superiority of haptic presentation. These experiments have both theoretical and practical implications. First, if it is the "haptic" nature of the stimulus movement that accounts for the results observed by Weisenberger and Hasser, theories of tactile stimulus processing would need to incorporate kinesthetic information about finger position and scanning velocity. In terms of development of tactile displays for telerobotic and virtual reality applications, knowledge of whether haptic movement of the observer was necessary for optimal perception would be important in display design.

In the first two experiments to be described, two possible explanatory factors were addressed. In Experiment 1, observers were permitted to repeat the stimulus as desired under three modes of stimulus presentation: static, passive scan, and haptic scan. In Experiment 2, observers were permitted, under the passive scan mode, to present the stimulus from multiple directions, to determine whether the presentation of different information in stimulus repetitions led to higher levels of performance than simple repetition of the same information.

A second series of experiments began investigation of a related question in tactile pattern perception. In these studies, the impact of reductions in the tactile "field of view" was evaluated. Loomis

(1980), as mentioned, found that a reduced-field slit-scan mode yielded higher levels of performance than the full-field mode when stimulus patterns were small and spatially dense. Loomis (1981) argued that the superiority of sequential presentation of stimuli might result from the limited spatial resolution of the tactile sense. If the tactile system acts as a low-pass filter for spatial sensing (relative to vision, for example), then a sequential scan of adjacent elements would give rise to a distribution of relative intensity of stimulation across the dimensions of the pattern. A simultaneous presentation of all elements in the pattern would yield a less accurate representation since the intensity of that pattern at each point is the convolution of the point-spread function at a given point with the entire input pattern, rather than the product of the pattern and the scanning aperture.

However, Loomis noted that the superiority of reduced-field presentation modes would be expected only for situations in which the spatial resolution abilities of the tactile system were challenged, i.e., for patterns complex enough to push the low-pass properties of the system. If so, then it must be assumed that the findings of a number of other studies (e.g., Craig, 1981), indicating that the sequential and slit-scan presentation modes were actually inferior to a full-field presentation, must have employed stimuli that did not challenge the spatial resolution abilities of the tactile system.

Results supporting Craig's (1981) finding were reported in Hill's (1974) study, in an additional experiment in which the number of columns of stimulation on the Optacon display was manipulated. On one fingertip, subjects were tested with 2, 4, 6, or 8 columns of simultaneous activation, displaying successively greater portions of an alphabetic pattern. Hill found a systematic increase in letter recognition accuracy as the number of columns was expanded from two to eight, suggesting that a larger field of view resulted in improved pattern recognition.

However, all of the above results were obtained with stationary tactile displays. That is, even though some of the presentation modes simulated pattern movement, the finger itself did not move. This is quite different from normal human sensing of object and surface properties, in which information is enhanced by movement of the finger relative to the surface being sensed. In fact, Lederman and Pawluk (1993) argue that in the absence of movement between skin and surface, perceptual information is essentially reduced to binary discriminations.

The question that arises is whether the deleterious effects of reducing the tactile field of view that were reported for stationary tactile displays can be generalized to a situation more similar to normal sensing, in which the finger moves relative to the pattern to be identified.

Given the possibility that pattern perception with active movement of the fingertip, under the control of the observer, might be qualitatively different in some respects from pattern recognition with stationary displays, the question arises whether reductions in the tactile field of view would yield deleterious effects similar to those observed by Craig (1981) and Hill (1974). To the degree that spatial extent is reduced on the display, the temporal aspects of the stimulus might be emphasized. The spatial

aspects of the stimulus would be increasingly dependent on the spatial feedback provided by kinesthetic cues from active movement. In the limit, if the field of view is reduced to a single element, the temporal aspects of the display are maximized.

There are a number of real-world tasks in which observers perform with a reduced tactile field of view. Katz (1925; 1989 translation) reported experiments in which individuals performed tactile tasks using an instrument to sense the surface, rather than by direct contact between skin and surface. In one study, subjects were able to identify the hardness of pencil leads by moving them across a piece of paper, based on vibrational cues transmitted to the fingers from friction between the pencil lead and the paper. In another study, subjects were able to discriminate the roughness of different grades of paper by moving a wooden rod across their surfaces. In these studies, surface features were sensed in an essentially temporal manner, since contact with the surface was a single point. Spatial cues were generated only kinesthetically. Results such as these suggest that when haptic movement is permitted, a reduced field of view might not be deleterious.

There is a strong, practically-motivated impetus for designing small and efficient tactile displays for use in virtual reality, telerobotic, and sensory aids applications. Displays with a smaller number of transducers would most likely be smaller in size, and might also impose less formidable processing and memory requirements, than larger displays. If haptic movement were a characteristic of the display situation, it is possible that a full-field display of actuators would not be required for accurate object and surface sensing. Results such as those of Katz support such a notion. However, the task used by Katz is very different from pattern identification.

In the present study, the question of whether haptic scanning of vibratory patterns is impaired by reductions in field of view was empirically addressed. In addition to practical considerations of display design, this question has theoretical implications for how pattern information is processed by the tactile system, particularly regarding the contribution of kinesthetic spatial information to pattern identification in situations where spatial cues generated by the display itself are reduced or absent. The first experiment employed a single-fingertip square display consisting of 9 elements, covering an area of 3.6 cm^2. Patterns used in the experiment were identical to those from an earlier study in which a full-fingertip, 30-element display (1.2 x 1.5cm) had been tested. The pattern sizes were chosen to fill this full-fingertip display, and thus were larger than the 9-element display used in the first experiment of the present study. Thus, the 9-element display constituted a reduced field of view for these patterns. The purpose of the first experiment was to determine whether these patterns could be identified with a reduced field-of-view display. Accordingly, results are considered in light of data reported for the same pattern sets by Weisenberger and Hasser (1994), in which a 30-element full-field display was used.

Experiment 1 -- Effects of multiple stimulus presentation on pattern identification

Method

Subjects. Four adults (3 males and 1 female, all nonsmokers), ranging in age from 19 to 23 years, were paid for their participation in the study. One subject (KJS) had participated in several earlier experiments on tactile pattern perception (reported in Weisenberger and Hasser, 1994); the remainder had no previous experience with tactile pattern perception. None reported any history of tactile dysfunction.

Apparatus. The shape-memory alloy (SMA) display used in the present study was designed and constructed by the TiNi Alloy Corporation (A. Johnson, 1991). The display contained 30 elements (5 columns by 6 rows) in a 1.5 cm x 1.2 cm array mounted in a plastic casing, such that the stimulator elements, when activated, protruded through holes in the casing. Center-to-center distance between elements was approximately 3 mm.

Each stimulator element was a beryllium-copper L-shaped cantilever beam. Anchored at each end of the beam was a thin titanium-nickel shape-memory alloy wire. When heated by input current, the SMA wire contracts, pulling the short end of the cantilever beam upwards; when current is discontinued, the wire cools and returns to its original configuration. Following an initial DC input to induce initial contraction, a pulse-width-modulated (PWM) current is input to the SMA wire to cause a vibratory output. Variations in pulse rate (frequency), duty cycle, and stimulus duration, as well as the duration of the initial DC pulse, are controllable in software. A more extensive characterization of the stimulator elements can be found in Hasser and Weisenberger (1993).

A CalComp Drawing Board II digitizing pad was employed for the haptic scan condition. The SMA display was mounted atop the pointer for the pad, and could be moved about the pad surface with preservation of the X-Y coordinates of the movement. The digitizing pad communicated with the computer via the mouse port.

The display and digitizing pad were interfaced to an 80386-based PC computer. The PC's serial port was used for communication with the display. C software permits specification of experimental runs, including values for each of the variables mentioned above, and controls stimulus presentation, response collection, and data storage. For the first experiment, duty cycle was fixed at 75% and the duration of the initial DC pulse was 100 ms.

The stimulus set consisted of 10 letters of the alphabet (A,K,X,B,R,G,C,D,O,Q), chosen based on data from Craig (1979) suggesting relatively high confusability of subsets of these letters. Stimuli were presented in the static mode, passive scan mode, and haptic scan mode. The static patterns were presented with a duration of 200 ms, and the passive scan patterns were presented with a duration of 100 ms per frame. Stimulus duration was not constrained in the haptic scan mode, and time in contact with the pattern was subject-determined. Three stimulus frequencies were tested: 20, 50, and 200 Hz, because Weisenberger and Hasser (1994) had found a significant advantage for low stimulus frequencies in their initial work.

For the static presentation mode, all pins making up a pattern were activated simultaneously, and continued to vibrate for the duration of the pattern. For the passive scan mode, patterns were constructed such that the left side of the pattern began on the right side of the finger, and successive frames of the stimulus moved the pattern in a leftward direction across the finger, in a manner analogous to that used by Craig (1981) for his scan mode. Offset of one frame of the stimulus was coincident with onset of the next frame. For the haptic scan mode, patterns were constructed by describing a set of points at a particular location on the digitizing tablet in X-Y space, which would cause activation when any portion of the SMA display reached the location of any point in the set. Movement of the digitizing tablet's pointer, with the SMA display mounted atop it, across the pad gave rise to the percept of a pattern in a fixed location on the pad.

Procedure. Testing was conducted in a quiet area of the laboratory, with the subject seated approximately 0.5 m from the PC monitor. The left index finger of the subject rested lightly on the plastic enclosure of the SMA display, and responses were entered on the keyboard of the PC with the right hand.

Stimulus presentation was arranged in 40-trial blocks. On each trial, a stimulus was presented, and subjects were prompted to choose the number corresponding to the icon of the presented pattern. Icons for each of the 10 patterns were displayed on the monitor throughout the block of trials. If the subject wished to repeat a stimulus, he pressed the letter "R" on the keyboard. After a response was entered, trial-by-trial feedback was provided on the monitor. In the haptic scan mode, subjects were permitted to scan in any direction for as long as desired before entering a response.

Within a 40-trial block, patterns were presented in a random order. For all trials in a block, values of stimulus frequency, stimulus duration, and presentation mode were fixed. Across blocks, stimulus frequency and presentation mode were varied randomly. Testing was conducted in sessions lasting 1-2 hours, with frequent rest periods to minimize fatigue. A total of 18 blocks was obtained from each subject.

Results and Discussion

Figure 1 shows the results, averaged across subjects, for pattern identification as a function of stimulus frequency, for the static, passive scan, and haptic scan presentation modes. As is evident from the figure, the ability to repeat the stimulus under all presentation modes had a strong effect on performance, effectively eliminating most of the differences observed among modes in Weisenberger and Hasser (1994). Unlike Weisenberger and Hasser's study, here it appeared that stimulus frequency was not an important variable, as performance levels did not vary across the frequencies tested.

A two-factor, within-subjects (presentation mode x frequency x subjects) analysis of variance was conducted on arcsine-transformed data to determine if there were any significant trends in the data. Results showed no differences that reached significance (mode of presentation, $F(2,6)=1.72$, not significant; stimulus frequency, $F(2,6)=1.02$, not significant; interaction of mode and frequency,

F(4,12)=0.73, not significant). These analyses confirmed the visual impression of Figure 1, that stimulus repetition eliminated differences among presentation modes.

However, there were differences in how the presentation modes were utilized by subjects. An examination of the number of repetitions selected for the static and passive scan presentation modes revealed that on average, subjects chose to repeat the stimulus more often in the static mode than in the passive scan mode (2.04 vs. 0.72 repeats, respectively). This difference was significant, as confirmed by a t-test (t(3)=10.5, p<.01). The fact that subjects needed more repetitions of the stimulus in the static mode to achieve comparable levels of performance to the passive scan mode suggests that they were able to glean more information from the stimulus in a single presentation in the passive scan mode, and thus that processing may have been different for the two modes of presentation.

Given the improvement in performance demonstrated for conditions in which the stimulus could be repeated, the question remained as to whether repetition of the same stimulus (and presumably the same information) provided as much benefit in pattern identification as would a condition in which each repetition provided different information, such as would be experienced if each repetition of the stimulus scanned the pattern from a different direction. Accordingly, in the second experiment, subjects were tested only in the passive scan mode, under several conditions: no repetition, repetition-same (each repetition presented the pattern scanned from the same direction), and repetition-different (each repetition presented the pattern scanned from a different direction).

Experiment 1 -- Effect of multiple scan directions

Method

Subjects. The four subjects from Experiment 1 also participated in Experiment 2.

Apparatus. General apparatus details are as in Experiment 1. As mentioned, only the passive scan mode was employed, under three presentation conditions: no repetition, repetition-same, and repetition-different. For all conditions, eight different scan directions were identified: top-bottom, bottom-top, right-left, left-right, top left-bottom right, bottom right-top left, top right-bottom left, and bottom left-top right. Under the no-repetition condition, for each 40-trial block a scan direction was selected, and all patterns were presented with that direction. Under the repetition-same condition, for each block a scan direction was selected, and all presentations, whether initial or repeated, had that direction. Under the repetition-different condition, for each block an initial scan direction was chosen, and the initial presentation of the stimulus had that direction. For any repetitions, subjects were free to choose any scan direction, with two caveats: they could not use the original direction again, and each repeat had to have a different direction. The number of repetitions was under subject control.

Procedure. General procedural details are as in Experiment 1. A total of 16, 40-trial blocks was obtained under each of the three presentation conditions, such that each initial scan direction was used twice under each condition. The order of presentation conditions, and the order of initial scan directions,

were randomized across blocks.

Results and Discussion

Figure 2 shows preliminary results, averaged across 3 subjects (the fourth subject was still being tested at the time of this report), for each presentation condition and each initial scan direction. These data indicate clearly that some scan directions yield higher levels of identification performance than others. Preferred directions appear to be the two horizontal scanning directions (right-left and left-right), and the four diagonal directions appear to be least useful in yielding stimulus information. It might have been expected that subjects would prefer direction 2 (right-left), since that was the default direction used in Experiment 1 for the passive scan mode, and the subjects had the most experience with that direction. However, performance for the other horizontal direction and for the two vertical directions was superior to that for the diagonal directions as well.

Closer examination of these data also reveals that for some scan directions, the use of different directions on stimulus repeats led to higher levels of performance. In general, it appeared that simple repetition of information from a non-preferred scan direction (e.g., the diagonal directions) was less beneficial than provision of new information from a preferred scan direction.

In Experiments 3 and 4, a different way of constraining stimulus presentation was investigated. Specifically, the effects of limiting the tactile "field of view" by employing displays that did not cover the entire fingertip was evaluated.

Experiment 3 -- Reducing the tactile "field of view"

Method

Subjects. Three adults (2 female, 1 male) between the ages of 23 and 38 participated in the experiment. All had some previous experience in tactile pattern perception tasks; CJH and JMW had participated in Weisenberger and Hasser's (1994) study, and KMS had participated in Experiments 1 and 2 of the present study. None reported any history of tactile dysfunction.

Apparatus. The display used in the present experiment was designed and constructed by the TiNi Alloy Corporation (A. Johnson, 1991). The display contained only nine actuators (3 columns by 3 rows) but was otherwise similar to the display described in Experiment 1. The stimulator elements, when activated, protruded through holes in the casing to contact the skin surface. Center-to-center distance between elements was approximately 3 mm, and the total area occupied by the stimulator array was 3.6 cm^2.

The actuators were constructed of a metal rod, which was connected to a thin length of titanium-nickel shape-memory alloy (SMA) wire. When heated by input current, the SMA wire contracts, forcing the rod upward; when current is discontinued, the wire cools and returns to its original configuration. Use of an input pulse-width-modulated current results in a vibratory output. Variations in pulse rate, frequency,

duty cycle, and stimulus duration, are controllable in software. Two sets of stimuli were tested in the first experiment. The first stimulus set, called "shapes," consisted of 8 simple geometric patterns constructed from one or two lines (-,\,X,I-,-I,+,/,I). The overall dimensions of these patterns were 1.2x1.2 cm; it should be noted that each of these patterns is larger than the stimulus display, such that an entire pattern could not be sensed at any point of contact. Rather, movement of the display was required for sensing of the entire pattern. This pattern set was employed because extensive earlier data had been obtained for these patterns with a similar display (Weisenberger & Hasser, 1994). The second set of patterns, called "letters," consisted of 10 letters used in Experiments 1 and 2. Several different vibration frequencies were tested in the present experiment. Frequencies were 10, 20, 50, and 200 Hz, and were generated by varying the pulse-width modulated current input to the actuators.

All patterns were presented under a haptic scan mode, in which the pattern was constructed as a set of points at a particular location on the digitizing tablet in X-Y space, which caused activation when any portion of the SMA display reached the location of any point in the set. Movement of the digitizing tablet's pointer, with the SMA display mounted atop it, across the pad gave rise to the percept of a pattern in a fixed location on the pad.

Procedure. General procedural details were identical to Experiments 1 and 2. Stimulus presentation was arranged in 40-trial blocks. On each trial, a stimulus was presented, and subjects were asked to move the SMA display into contact with the pattern. Subjects were permitted to scan the pattern in any direction for as long as desired before entering a response. After completion of scanning, subjects were prompted to choose the number corresponding to the icon of the pattern. Icons for each of the patterns were displayed on the monitor throughout the block of trials. After a response was entered, trial-by-trial feedback was provided on the monitor.

Within a block, patterns were presented in a random order. For all trials in a block, values of stimulus frequency were fixed. Across blocks, stimulus frequency was varied randomly. All blocks employing the set of shapes were presented first, followed by all blocks employing the set of letters, in order to avoid negative transfer effects between sets (in our experience, subjects found the set of letters more difficult, and performed more poorly, if they had just completed a block of shapes than if they had completed another block of letters). Testing was conducted in sessions lasting 1-2 hours, with frequent rest periods to minimize fatigue. A total of 16 blocks of data was obtained from each subject.

Results and Discussion

Figure 3 shows averaged data from all subjects for pattern identification performance as a function of frequency, for each of the pattern sets. As can be seen, performance with the set of shapes was high, averaging well over 90 percent correct for all presentation frequencies. Performance with the set of letters was consistently lower, ranging from 76-83 percent, with a mean of 79.5 percent. These results are consistent with reports from Weisenberger and Hasser (1994) for the same pattern sets; they also found

that the set of letters produced consistently lower performance than the set of shapes.

In fact, the data obtained in the present experiment are quite comparable to those of Weisenberger and Hasser's study. In that study, overall performance levels (averaged across subjects and stimulus frequencies) of approximately 90 percent correct were found for the set of shapes, and levels of 65-75 percent correct were found for the set of letters. These numbers correspond well to the overall levels of 94 percent and 79 percent for the sets of shapes and letters in the present experiment. This comparability is interesting given that the display used by Weisenberger and Hasser (1994) was a full-field, 30-element display, and that the patterns used in both experiments were sized to fit this 30-element display. Thus, the results suggest that reduction of the field of view for tactile pattern stimulation did not impair pattern recognition (if anything, performance was somewhat better with the 9-element display). It should be noted, however, that the 9-element display produced a much more consistent and robust vibratory percept than did the 30-element display. Inconsistencies in stimulator fabrication for the 30-element display led to inconsistent stimulator activation, both in rise time and in extent of excursion of the actuator. These differences may have contributed to the relatively good performance observed with the 9-element display.

A closer look at the data from the present experiment reveals two trends that also parallel findings from Weisenberger and Hasser (1994). First, performance for the set of shapes was much better than that for the set of letters, confirming the notion that patterns in the set of letters are much more difficult to discriminate. Second, there is some suggestion of lower levels of performance for higher-frequency stimulus presentation conditions, which is consistent with the idea that the tactile mechanoreceptors most responsive to fine spatial detail operate most effectively with low-frequency stimulation (Johnson and Hsiao, 1991).

Analysis of confusion matrices from the present experiment also yielded interesting findings. There were few confusions of any sort for the set of shapes, as would be expected given the high overall percent correct levels. Confusions for the set of shapes were quite orderly, with confused items primarily ones that contained similar features. For example, pattern 2, an "X", was confused primarily with patterns 1 and 6, each of which contained one diagonal line. This pattern was not frequently confused with patterns that did not contain diagonals. Confusion matrices for the set of letters are more widespread. Some general characteristics are discernible for these data as well. First, patterns corresponding to "G" and "Q" were frequently confused, as were "R" and "K", and "D" and "O." This ordering of confusion data was highly similar to confusion matrices obtained by Weisenberger and Hasser (1994) for the 30-element display. The similarity in confusion matrices for the two displays supports the notion that the processing done by observers using the 9-element display was similar to that for observers using the 30-element display. That is, the fact that the 9-element display permits only partial access to pattern elements at any point in time did not induce differences in perceived pattern features.

The finding that performance for the 9-element display was so similar to that for the 30-element display argues that observers can construct a coherent pattern from a number of successively-experienced spatial images. In other words, the process of pattern perception, at least for the 9-element display, of necessity had a temporal component. The question naturally arises as to whether further reductions in the tactile field of view, which would introduce an even greater temporal component into the pattern perception process, would have an impact on performance. Accordingly, in the fourth experiment, the 9-element display was compared with two additional displays, one with 4 elements, and one with 1 element. The 1-element display presents an interesting situation to the observer, in that for this condition, there is only one location on the skin surface that receives stimulation, and the spatial nature of the percept is constructed entirely from kinesthetic feedback resulting from active movement of the fingertip relative to the pattern to be sensed. As mentioned, there is also a practical consideration in determining just how far the tactile field of view can be reduced without deleterious effects on pattern identification, given the space and technological sophistication required to construct wearable, movable tactile displays that might be mounted in data gloves for teleoperation and virtual sensing applications.

Experiment 4 - Further reductions in the tactile field of view

Method

Subjects. Three subjects (two female, one male) participated. The two female subjects had participated in Experiment 3, whereas the male subject had participated in previous studies involving the 30-element version of the display. All subjects were practiced until they had reached stable levels of performance on all presentation conditions before actual data collection was initiated.

Apparatus. The basic apparatus was identical to that for Experiment 3. The 9-element display was used in all presentation modes. The three different presentation modes employed in the present experiment were achieved by activating only a subset of elements on the display for the 4-element and 1-element display modes. For the 4-element display, four adjacent elements in the upper right corner of the 9-element display were activated (rows 1 and 2 of the rightmost column, and rows 1 and 2 of the center column). For the 1-element display, the center element of the 9-element display was activated. Selective activation of subsets of elements was accomplished in software.

Stimuli again consisted of the sets of shapes and letters used in Experiment 3. No modifications to the stimuli themselves were performed.

Procedure. General procedural details were identical to Experiment 3. Three presentation modes were tested: 9-element, 4-element, and 1-element displays. Patterns were presented in random order in 40-trial blocks. All blocks of patterns from the set of shapes were completed before testing with the set of letters was begun. All patterns within a block were presented in the same presentation mode. The three presentation modes (9-element, 4-element, and 1-element) were tested in random order across blocks. Each mode was tested three times for each set of patterns, for a total of 120 trials per subject for each

mode and each pattern set. For simplicity, all stimuli were presented at a frequency of 20 Hz.

<u>Results and Discussion</u>

Figure 5 shows averaged performance for two of the three subjects (the third subject had not completed testing at the time of this report) for each of the three presentation modes: 9-element, 4-element, and 1-element, for the set of shapes and the set of letters. Several aspects of these results are immediately apparent. First, as before, overall levels of performance are higher for the set of shapes than for the set of letters, confirming that the set of shapes contains patterns that are easier to identify. Second, and more interestingly, performance levels for the 4-element display are not substantially reduced relative to those for the 9-element display. However, it is evident that reducing the spatial extent of the display to a single element has an extremely serious consequence, since performance is reduced to almost chance levels for the 1-element display.

This finding suggests that there are indeed limits on reducing the tactile field of view, if performance levels are to be maintained. The fact that performance levels are at all above chance suggests that some information can be obtained from the single-element display, but the loss of spatial extent on the display results in a substantial reduction of available information. On the other hand, the reduction from 9 elements to 4 elements did not have an appreciable effect on performance, suggesting that the tactile system can make do with input from a rather small spatial area on the fingertip, as long as movement relative to the stimulus is permitted.

Confusion matrices, averaged across subjects, for the 9-element, 4-element, and 1-element presentation modes for the set of shapes are shown in Figures 6-8. Comparison of the confusions generated for the 9-element and 4-element displays, in Figures 6 and 7, indicate substantial similarities in processing of the patterns presented under these two display modes. However, confusions for the 1-element display are less similar, suggesting that subjects may have pursued alternative processing strategies for this presentation mode.

It is also enlightening to examine differences in scanning behavior across the 9-element, 4-element, and 1-element display modes. Differences in duration of pattern exploration were measured for one subject (KMS) for each of the presentation modes. A total of 10 stimulus presentations were timed for each presentation mode, and averages for each mode were obtained. For this subject, the average scanning time for the 9-element display, from the time the stimulus was first contacted till the response was completed, was 9.1 s. For the 4-element display, this time was appreciably longer, averaging 15.8 s. For the 1-element display, scan duration was longer still, averaging 24.2 s. Thus, it can be seen that although performance levels were not decreased when display size was reduced from 9 element to 4 elements, nonetheless there was an increase in the amount of time required for pattern identification, suggesting that the two presentation modes were not equally easy to use. For the 1-element display, subjects were not able to use scanning duration as a tradeoff for accuracy--even with a 24 s scanning

duration, performance levels were still extremely low.

Table 1 -- Scan durations for different display sizes (subject KMS)

9-element	4-element	1-element
9.1 ms	15.8 ms	24.2 ms

Overall, the results of Experiment 4 indicate that the tactile field of view can indeed be reduced considerably in spatial extent, without substantial impact on performance. However, this reduction cannot be extended all the way down to a single element without loss of information. Although subjects were able to extract some information from the 1-element display, performance did not approach levels for the 4-element or 9-element displays. Further, differences in scanning time across the presentation modes have implications for the design of practically useful displays, in that for some applications the user may not have unlimited time to access the tactile stimulus.

Loomis (1981) suggested that there might be an interaction between display size and pattern size, such that patterns with substantial spatial detail (e.g., very small patterns) might be more easily sensed through a small display, whereas larger patterns might be less affected by the size of the display. The basis of this argument was Loomis' suggestion that the tactile system acts as a low-pass spatial filter for pattern recognition. According to his argument, larger patterns would not tax the spatial resolution abilities of the system, since they would contain primarily lower spatial frequencies. However, if the same patterns were reduced in size, they would of necessity have a greater proportion of high-spatial frequency information. These patterns would be more easily sensed by a smaller display than a larger display, since the smaller display would limit the diffusion introduced by low-pass filtering of the tactile system. In Experiments 3 and 4 of the present study, pattern size was held constant, and the patterns were larger than any of the displays tested (9, 4, or 1 element). It is possible that the patterns used in these experiments did not significantly tax the low-frequency filtering characteristics of the tactile system, and thus no differences of display size were seen. For a pattern set that did contain a high proportion of high-frequency information, Loomis' argument suggests that the 4-element display would be superior to the 9 element display, and thus the notion is that reducing the tactile field of view might actually enhance performance. This notion was not supported by the results of Experiment 4, but work with stimuli of different sizes might shed additional light on this discrepancy. Further experiments are planned to address this question.

General conclusions

Overall, the results of the four experiments of the present study add important information to our

understanding of how complex tactile pattern stimuli are processed by human observers. These data have implications for the design and implementation of tactile displays for virtual sensing and teleoperation. First, it appears that the ability to repeat stimuli is an important component of haptic perception, and that haptic capability may not be absolutely necessary for some surface identification tasks, as long as stimuli can be sensed repeatedly. Further, the ability to present different information on subsequent presentations of the stimulus, as in Experiment 2, seems to provide additional benefits to performance, and suggests that displays that could simulate the percept of scanning of a surface from different directions may be superior to ones that permit only a single direction of scanning. Experiments 3 and 4 suggest that tactile displays for virtual sensing or telerobotics need not cover the entire surface of the fingertip. For many tasks, a smaller display may yield performance that is as good as that for a full-fingertip display. For implementation and packaging of displays, the ability to reduce display size and the number of display elements provides a simplification that may make such implementations more feasible. Finally, the versatility of the display used in the present study makes it an extremely useful tool for basic investigations of tactile perception. The ability to vary stimulus frequency and duty cycle, together with the ability to simulate haptic sensing of surface features, permit investigation of questions of tactile processing that have not previously been amenable to study. Further work with this display should facilitate understanding of the processes involved in tactile pattern perception.

REFERENCES

Beauchamp, K.L., Matheson, D.W., and Scadden, L.A. (1971). Effects of stimulus-change method on tactile-image recognition. Perceptual and Motor Skills, 33, 1067-1070.

Craig, J.C. (1976). Vibrotactile letter recognition: The effects of a masking stimulus. Perception & Psychophysics, 20, 317-326.

Craig, J.C. (1977) Vibrotactile pattern perception: Extraordinary observers. Science, 196, 450-452.

Craig, J.C. (1978). Vibrotactile pattern recognition and masking. In: Active Touch--The Mechanisms of Recognition of Objects by Manipulation: A Multidisciplinary Approach. Ed. G. Gordon (Oxford: Pergamon Press).

Craig, J.C. (1979). A confusion matrix for tactually presented letters. Perception & Psychophysics, 15, 123-127.

Craig, J.C. (1980). Modes of vibrotactile pattern generation. Journal of Experimental Psychology: Human Perception and Performance, 6, 151-166.

Craig, J.C. (1981). Tactile letter recognition: Pattern duration and modes of pattern generation. Perception & Psychophysics, 31, 540-546.

Craig, J.C. (1982). Vibrotactile masking: A comparison of energy and pattern maskers. Perception & Psychophysics, 31, 523-529.

Craig, J.C. (1983). The role of onset in the perception of sequentially presented vibrotactile patterns. Perception & Psychophysics, 34, 421-432.

Craig, J.C. (1985). Attending to two fingers: Two hands are better than one. Perception & Psychophysics, 38, 496-511.

Craig, J.C. and Evans, P.M. (1987). Vibrotactile masking and the persistence of tactual features. Perception & Psychophysics, 42, 309-317.

Evans, P.M. and Craig, J.C. (1986). Temporal integration and vibrotactile backward masking. Journal of Experimental Psychology: Human Perception & Performance, 12, 160-168.

Hasser, C.J. and Weisenberger, J.M. (1993). Preliminary evaluation of a shape-memory-alloy tactile feedback display. ASME Advances in Robotics, Mechatronics, and Haptic Interfaces, DSC-49, 73-80.

Hill, J.W. (1974). Limited field of view in reading lettershapes with the finger. In F.A. Geldard (Ed.), Cutaneous communication systems and devices. Austin, TX: The Psychonomic Society.

Johansson, R.S. and Westling, G. (1987). Signals in tactile afferents from the fingers eliciting adaptive motor responses during precision grip. Experimental Brain Research, 66, 141-154.

Johnson, A.D. (1990). Shape-memory alloy tactile feedback actuator. Technical Report AAMRL- TR-90-039. TiNi Alloy Company, 1144 65th Street, Oakland CA, August.

Katz, D. (1989). The World of Touch (L.E. Krueger, Trans). Hillsdale, N.J.: Lawrence Erlbaum Assoc. (Original work published 1925).

Kirman, J.H. (1974). Tactile apparent motion: the effect of interstimulus onset interval and stimulus duration. Perception and Psychophysics, 15, 1-6.

Lederman, S.J. and Pawluk, D.T. (1993). Lessons from the study of biological touch for robot tactile sensing. In: Advanced Tactile Sensing for Robotics. Ed. H. Nicholls, World Press: London, 1-44.

Loomis, J.M. (1974). Tactile letter recognition under different modes of stimulus presentation. Perception and Psychophysics, 16, 401-408.

Loomis, J.M. (1980). Interaction of display mode and character size in vibrotactile letter recognition. Bulletin of the Psychonomic Society, 16, 385-387.

Loomis, J.M. (1981). Tactile pattern perception. Perception, 10, 5-27.

Sherrick, C.E. and Rogers, R. (1966). Apparent haptic movement. Perception and Psychophysics, 1, 175-180.

Weisenberger, J.M., & Hasser, C.J. (1994). Role of active and passive movement in vibrotactile pattern perception. Submitted for publication.

White, B.W., Saunders, F.A., Scadden, L., Bach-y-Rita, P., and Collins, C.C. (1970). Seeing with the skin. Perception and Psychophysics, 7, 23-27.

Figure 1. Percent correct identification performance as a function of frequency for static, passive scan, and haptic scan presentation modes. Data are averaged for three subjects.
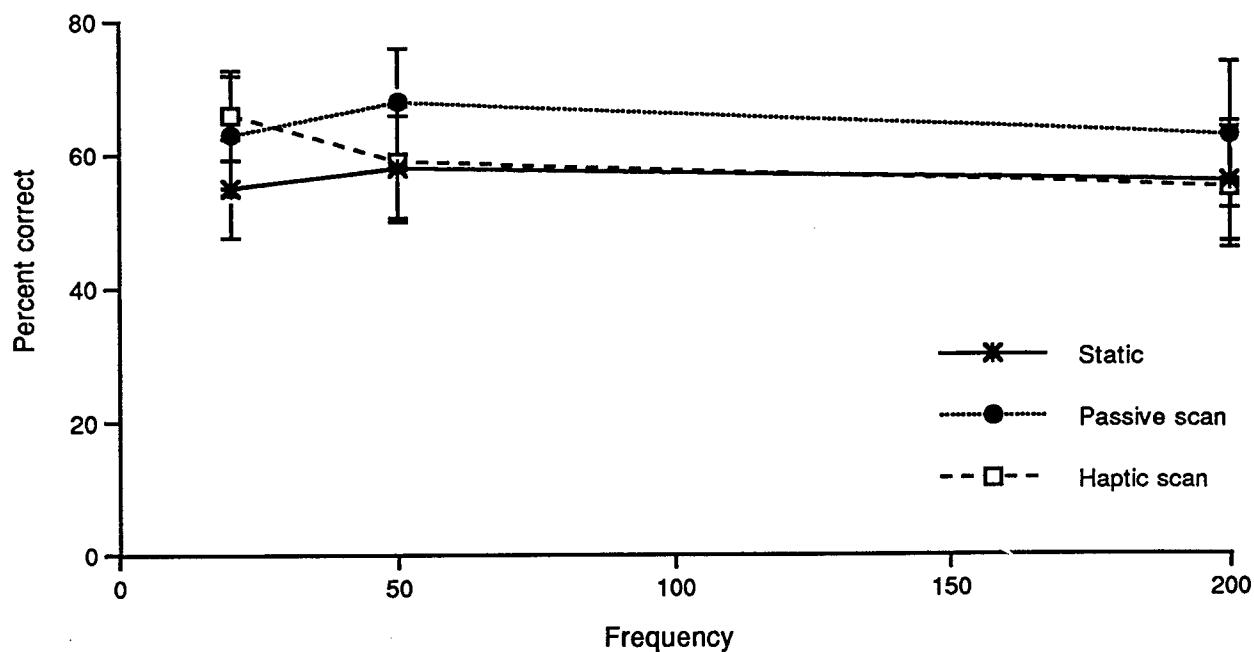

Figure2. Percent correct letter identification as a function of scan direction, for no repetition, repetition-same, and repetition-different presentation conditions. Data are averaged for three subjects.

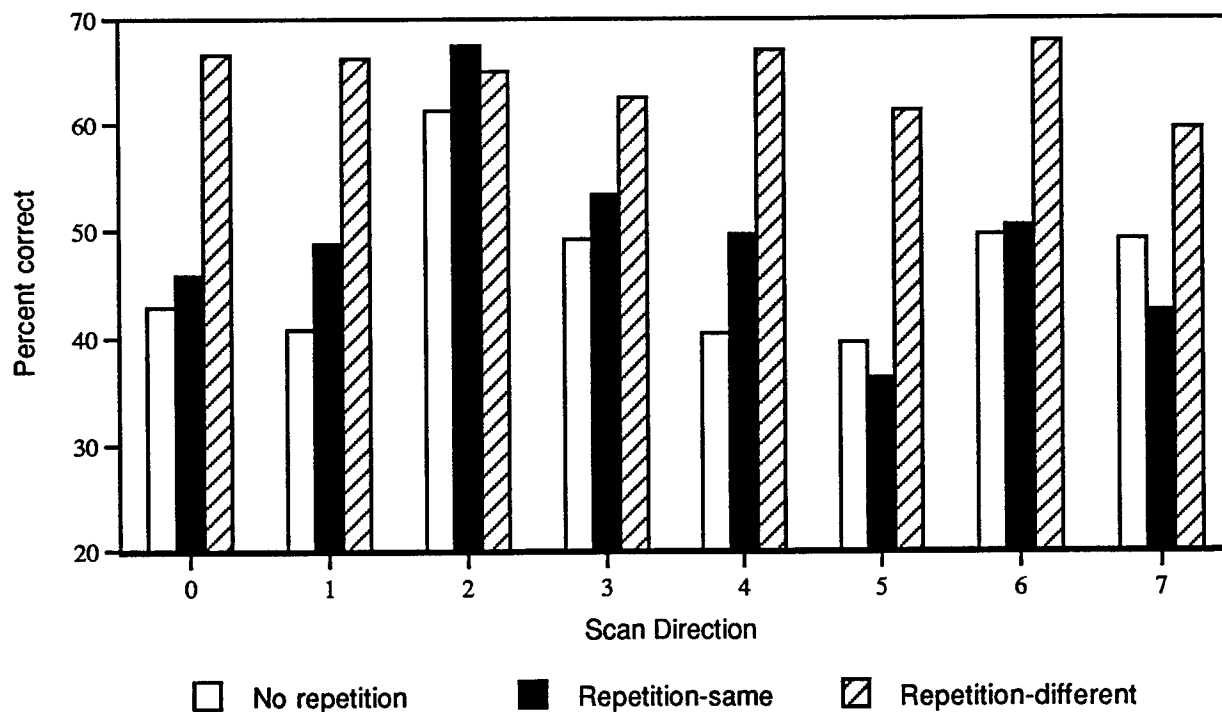☐ No repetition   ■ Repetition-same   ▨ Repetition-different

45-19

Figure3. Percent correct pattern identification as a function of frequency for the sets of shapes and letters. Data are averaged for three subjects.
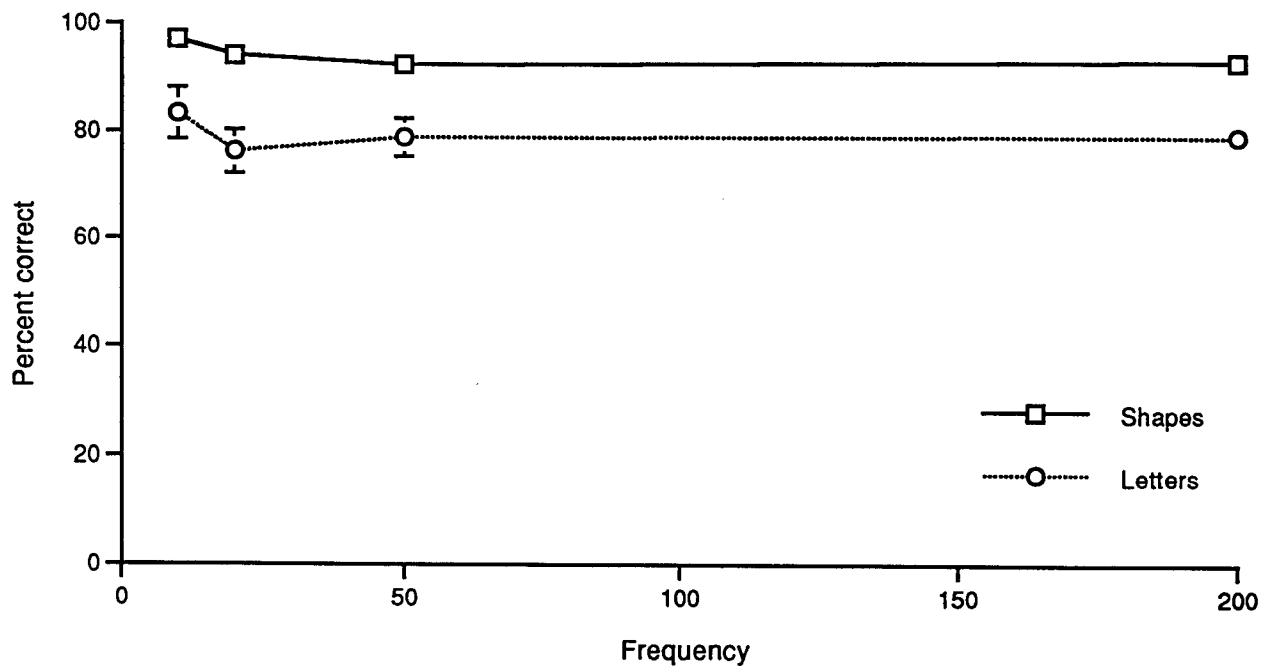

Figure 4. Percent correct identification performance for the 1-element, 4-element, and 9-element displays, for the sets of shapes and letters. Data are averaged for two subjects.